Top 10 Words Per Item (index-based labels)

=== item_0 ===

rule: 4, companies: 3, regulation: 3, americans: 3, securities: 2, exchange: 2, commission: 2, sec: 2, force: 2, disclose: 2

=== item_1 ===

fed: 4, treasury: 4, two: 3, financial: 3, federal: 2, made: 2, department: 2, would: 2, trillion: 2, bonds: 2

=== item_2 ===

traffic: 4, virginia: 4, stuck: 3, leesburg: 3, pike: 3, northern: 3, jam: 3, road: 3, anyone: 3, said: 3

=== item_3 ===

drones: 4, government: 4, use: 3, said: 3, coalition: 2, allow: 2, kill: 2, terrorists: 2, letter: 2, continuing: 2

=== item_4 ===

said: 4, software: 3, old: 2, still: 2, many: 2, schulz: 2, cylance: 2, company: 2, technology: 2, time: 2

=== item_5 ===

marijuana: 6, measure: 4, law: 3, would: 3, federal: 3, one: 2, long: 2, state: 2, allow: 2, possess: 2

=== item_6 ===

consumers: 6, banks: 4, financial: 3, cfpb: 3, debt: 3, urged: 2, consumer: 2, protection: 2, misleading: 2, payday: 2

=== item_7 ===

encryption: 6, debate: 2, export: 2, sens: 2, believe: 2, government: 2, allowed: 2, name: 2, national: 2, security: 2

=== item_8 ===

companies: 6, government: 5, devices: 4, bill: 4, new: 3, would: 3, draft: 2, unlock: 2, court: 2, compel: 2

=== item_9 ===

financial: 3, new: 3, futures: 3, trading: 3, commission: 3, cftc: 3, institutions: 2, transparency: 2, industry: 2, commodity: 2

=== item_10 ===

central: 4, bank: 4, ruble: 2, russian: 2, percent: 2, tapering: 2, effect: 2, markets: 1, across: 1, eastern: 1

=== item_11 ===

percent: 4, index: 3, rose: 3, stock: 2, exchange: 2, prague: 1, continued: 1, gain: 1, ground: 1, thursday: 1

=== item_12 ===

stock: 4, percent: 4, czech: 3, government: 3, reforms: 3, exchange: 2, index: 2, nearly: 2, last: 2, months: 2

=== item_13 ===

percent: 6, stock: 5, close: 4, bulgaria: 4, market: 3, index: 3, central: 3, bank: 3, thursday: 2, hit: 2

=== item_14 ===

central: 6, bank: 6, czech: 5, inflation: 4, percent: 4, rate: 3, interest: 3, november: 2, month: 2, european: 2

=== item_15 ===

czech: 8, parliament: 5, bill: 4, republic: 4, seats: 3, lower: 3, house: 3, klaus: 2, support: 2, reduce: 2

=== item_16 ===

iran: 6, deal: 3, russia: 3, nuclear: 3, talks: 3, us: 3, year: 3, would: 3, czech: 2, klaus: 2

=== item_17 ===

czech: 1, economic: 1, growth: 1, expected: 1, continue: 1, slowing: 1, next: 1, years: 1, country: 1, get: 1

=== item_18 ===

country: 4, czech: 4, markets: 3, banking: 3, bank: 3, position: 2, week: 2, already: 2, system: 2, republic: 2

=== item_19 ===

song: 10, czech: 9, folk: 4, republic: 2, eurovision: 2, roma: 2, good: 2, example: 2, music: 2, traditional: 2

=== item_20 ===

year: 4, sold: 4, property: 3, last: 3, england: 3, wales: 3, homes: 3, market: 2, report: 2, also: 2

=== item_21 ===

grenfell: 6, deutsche: 5, tower: 5, bank: 5, firm: 4, letter: 3, london: 3, stock: 3, exchange: 3, ritchie: 2

=== item_22 ===

exchange: 6, foreign: 5, trading: 4, government: 4, defend: 3, treasury: 3, budget: 3, also: 3, bank: 3, chancellor: 3

=== item_23 ===

fund: 5, horlick: 3, clark: 3, superwoman: 2, thursday: 2, years: 2, based: 2, million: 2, story: 2, continues: 2

=== item_24 ===

fund: 5, pension: 4, board: 4, horlick: 3, superwoman: 2, manager: 1, nicola: 1, announced: 1, resignation: 1, tuesday: 1

=== item_25 ===

grenfell: 3, former: 2, nicola: 2, horlick: 2, told: 2, judge: 2, victims: 2, family: 2, money: 2, kensington: 2

=== item_26 ===

deutsche: 3, grenfell: 3, building: 3, fire: 3, bank: 3, horlick: 2, job: 2, safety: 2, statement: 2, said: 2

=== item_27 ===

grenfell: 7, bank: 4, tower: 4, fire: 4, council: 3, financial: 3, greste: 3, investment: 2, deutsche: 2, times: 2

=== item_28 ===

bank: 5, said: 4, country: 4, germany: 3, deal: 3, japan: 3, government: 3, agreement: 3, allow: 2, deutsche: 2

=== item_29 ===

merrill: 6, lynch: 6, investment: 4, banking: 4, tallahassee: 4, firm: 3, operations: 2, state: 2, office: 2, coconut: 2

=== item_30 ===

years: 6, liu: 5, xiaobo: 5, state: 3, power: 3, chinese: 2, dissident: 2, said: 2, whose: 2, name: 2

=== item_31 ===

china: 5, kong: 5, hong: 4, announcement: 3, chinese: 3, government: 3, thursday: 2, territory: 2, autonomous: 2, propaganda: 1

=== item_32 ===

said: 5, china: 4, international: 4, party: 3, jiang: 3, role: 3, system: 3, way: 3, communist: 2, prepared: 2

=== item_33 ===

help: 6, china: 4, companies: 4, pledged: 3, foreign: 3, economic: 3, government: 3, also: 3, investment: 2, invest: 2

=== item_34 ===

china: 4, said: 3, bank: 3, yuan: 3, dollar: 3, borrowing: 2, costs: 2, people: 2, percent: 2, roll: 2

=== item_35 ===

said: 6, tighten: 5, goods: 4, controls: 4, korea: 4, china: 3, imports: 3, ministry: 3, sunday: 2, beijing: 2

=== item_36 ===

yuan: 6, china: 5, said: 3, demand: 3, record: 3, dollar: 3, overheating: 2, strength: 2, chinese: 2, value: 2

=== item_37 ===

government: 8, plans: 3, education: 3, development: 3, retail: 2, reduce: 2, cost: 2, number: 2, working: 2, sector: 2

=== item_38 ===

wang: 7, chinese: 4, american: 4, group: 3, china: 3, detained: 2, donghai: 2, president: 2, political: 2, times: 2

=== item_39 ===

activists: 2, names: 2, china: 1, jailed: 1, dozens: 1, dissidents: 1, year: 1, amid: 1, beijing: 1, campaign: 1

=== item_40 ===

digital: 6, quarter: 5, ad: 5, company: 4, mobile: 4, media: 3, revenue: 3, higher: 3, said: 3, driven: 3

=== item_41 ===

bank: 8, lending: 4, billion: 4, commonwealth: 3, australia: 3, home: 3, business: 3, mortgage: 3, loan: 2, posted: 2

=== item_42 ===

move: 7, bank: 5, dollar: 5, currency: 5, australian: 4, would: 4, time: 4, first: 3, major: 3, australia: 2

=== item_43 ===

percent: 7, dollar: 7, peso: 5, since: 4, greenback: 3, wednesday: 2, would: 2, government: 2, record: 2, high: 2

=== item_44 ===

bank: 8, strong: 3, grow: 3, year: 2, good: 2, retail: 2, said: 2, areas: 2, business: 2, commonwealth: 1

=== item_45 ===

said: 4, newspaper: 3, magazine: 3, radio: 2, suspended: 2, deeply: 2, saddened: 2, news: 2, necessary: 2, ensure: 2

=== item_46 ===

lending: 6, bank: 5, australia: 5, said: 5, program: 4, competitive: 4, murray: 3, spain: 3, market: 3, never: 2

=== item_47 ===

seven: 10, network: 7, news: 6, west: 4, media: 4, channel: 3, chief: 3, deal: 2, business: 2, new: 2

=== item_48 ===

company: 3, brambles: 2, shipment: 2, south: 2, east: 2, asia: 2, delivering: 2, expanded: 2, pallet: 1, hire: 1

=== item_49 ===

stadium: 6, olympic: 5, built: 4, consortium: 3, billion: 3, sports: 3, venue: 3, cost: 3, facility: 3, melbourne: 2

=== item_50 ===

company: 6, said: 4, sale: 3, friday: 2, looking: 2, sell: 2, ready: 2, make: 2, final: 2, decision: 2

=== item_51 ===

finance: 4, fitch: 4, mercury: 3, arizona: 3, consumer: 2, company: 2, pay: 2, said: 2, ratings: 2, case: 2

=== item_52 ===

company: 7, said: 3, quarter: 3, loss: 3, forecast: 2, business: 2, financial: 2, creditors: 2, story: 2, continues: 2

=== item_53 ===

san: 4, francisco: 4, fbi: 3, said: 3, charges: 3, office: 3, fraud: 3, mercury: 2, company: 2, federal: 2

=== item_54 ===

said: 5, focus: 4, business: 4, chief: 3, executive: 3, core: 3, brown: 3, bank: 3, unit: 3, canadian: 3

=== item_55 ===

p: 3, dow: 3, nasdaq: 3, strong: 2, market: 2, months: 2, recent: 2, doubled: 2, composite: 2, business: 1

=== item_56 ===

khosla: 5, india: 3, ibi: 3, business: 2, forbes: 2, first: 2, two: 2, group: 2, indian: 2, company: 2

=== item_57 ===

said: 5, thursday: 3, suspension: 3, aware: 3, finance: 2, contract: 2, agency: 2, cra: 2, mercury: 1, suspending: 1

=== item_58 ===

company: 4, california: 4, said: 3, state: 3, tax: 3, mercury: 2, finance: 2, million: 2, royalties: 2, federal: 2

=== item_59 ===

company: 5, said: 4, india: 3, reliance: 2, debt: 2, march: 2, taken: 2, sector: 2, acceptance: 1, corp: 1

=== item_60 ===

said: 4, merger: 3, gold: 3, would: 2, firms: 2, deal: 2, create: 2, copper: 2, producer: 2, canada: 2

=== item_61 ===

police: 4, said: 4, man: 3, think: 3, officer: 3, thursday: 2, officers: 2, home: 2, found: 2, halifax: 2

=== item_62 ===

gas: 5, energy: 3, increase: 3, amount: 3, pipeline: 3, system: 3, supply: 2, moore: 2, province: 2, ontario: 2

=== item_63 ===

city: 7, residents: 5, new: 4, scarborough: 3, hotel: 2, northern: 2, massive: 2, billion: 2, coming: 2, safety: 2

=== item_64 ===

mayor: 5, city: 4, ford: 4, income: 4, toronto: 3, allegations: 3, investigation: 3, brother: 2, doug: 2, facing: 2

=== item_65 ===

gas: 7, indonesian: 3, said: 3, mining: 2, country: 2, made: 2, progress: 2, new: 2, partner: 2, mine: 2

=== item_66 ===

annual: 2, molson: 1, cos: 1, ltd: 1, hold: 1, shareholder: 1, meeting: 1, april: 1, new: 1, york: 1

=== item_67 ===

police: 6, said: 5, two: 5, charges: 3, men: 3, one: 3, laid: 2, woman: 2, arrested: 2, package: 2

=== item_68 ===

cover: 7, often: 5, wear: 5, outfits: 5, way: 5, stage: 5, dance: 3, naked: 3, dancers: 3, clubs: 3

=== item_69 ===

de: 5, guzman: 5, michael: 3, family: 3, death: 2, geologist: 2, oil: 2, discovery: 2, thursday: 2, said: 2

=== item_70 ===

bolt: 6, ev: 6, miles: 3, charge: 3, offering: 2, consumers: 2, new: 2, offer: 2, models: 2, range: 2

=== item_71 ===

union: 7, first: 5, auto: 4, target: 4, said: 4, workers: 3, chrysler: 3, canada: 3, cwa: 3, canadian: 2

=== item_72 ===

quarter: 7, said: 6, profit: 6, second: 4, share: 3, expects: 2, gm: 2, gross: 2, automaker: 2, sales: 2

=== item_73 ===

photos: 16, fire: 9, south: 8, gunman: 8, opens: 8, church: 8, charleston: 8, carolina: 7, shooting: 2, missouri: 2

=== item_74 ===

breakfast: 5, new: 5, lunch: 4, twin: 4, cities: 4, paul: 4, kellogg: 3, food: 3, buy: 2, create: 2

=== item_75 ===

breakfast: 8, kellogg: 3, food: 3, offerings: 3, billion: 3, move: 2, broaden: 2, past: 2, quaker: 2, oats: 2

=== item_76 ===

workers: 7, deal: 4, united: 3, auto: 3, union: 3, gm: 3, ratified: 2, new: 2, labour: 2, contract: 2

=== item_77 ===

retail: 3, retailer: 3, going: 2, model: 2, choice: 2, amazon: 2, stores: 2, revenue: 2, growth: 2, using: 1

=== item_78 ===

cost: 2, know: 2, impact: 2, going: 2, store: 2, small: 2, using: 1, cuts: 1, offset: 1, flat: 1

=== item_79 ===

gm: 7, chrysler: 6, would: 4, shares: 3, bankruptcy: 3, rescue: 2, financial: 2, percent: 2, surprised: 2, go: 2

=== item_80 ===

said: 3, parts: 3, talks: 2, company: 2, military: 2, journal: 2, contract: 2, exploring: 2, aerospace: 1, giant: 1

=== item_81 ===

mobile: 5, itu: 5, market: 5, sales: 4, phone: 4, said: 4, forecast: 3, global: 3, expected: 3, latest: 2

=== item_82 ===

aircraft: 6, saudi: 5, yemen: 5, coalition: 4, hit: 3, source: 3, group: 2, arabia: 2, crossed: 2, border: 2

=== item_83 ===

said: 4, company: 4, purchase: 3, brown: 3, close: 3, factory: 2, could: 2, significant: 2, impact: 2, plant: 2

=== item_84 ===

yen: 8, profit: 6, billion: 4, percent: 4, net: 4, company: 4, year: 3, hanson: 2, reported: 2, increase: 2

=== item_85 ===

business: 3, hanson: 2, billion: 2, company: 2, share: 2, price: 2, july: 2, conglomerate: 1, plc: 1, reported: 1

=== item_86 ===

bombing: 6, syria: 5, syrian: 5, government: 5, united: 4, states: 4, britain: 3, today: 2, end: 2, campaign: 2

=== item_87 ===

aleppo: 7, united: 3, nations: 3, russia: 3, said: 3, british: 2, government: 2, friday: 2, situation: 2, humanitarian: 2

=== item_88 ===

swedish: 7, government: 6, said: 5, court: 4, decision: 4, would: 3, legal: 3, british: 2, european: 2, justice: 2

=== item_89 ===

ban: 5, eu: 4, british: 3, european: 3, commission: 3, travel: 3, said: 3, government: 2, could: 2, people: 2

=== item_90 ===

company: 9, lower: 7, flat: 6, revenues: 6, expects: 5, full: 5, year: 5, second: 4, quarter: 3, operating: 3

=== item_91 ===

vlsi: 4, stock: 3, russia: 3, lukoil: 3, company: 3, investors: 3, technology: 2, friday: 2, closing: 2, giant: 2

=== item_92 ===

president: 6, know: 5, think: 4, statement: 4, bold: 4, position: 3, paragraph: 2, torture: 2, administration: 2, first: 2

=== item_93 ===

percent: 6, shares: 4, trading: 4, company: 3, extended: 3, expected: 3, first: 2, said: 2, stock: 2, letter: 2

=== item_94 ===

said: 2, china: 2, india: 2, computer: 1, associates: 1, international: 1, inc: 1, tuesday: 1, north: 1, america: 1

=== item_95 ===

game: 5, decision: 3, frozen: 3, computer: 2, associates: 2, international: 2, company: 2, said: 2, suspend: 2, sale: 2

=== item_96 ===

core: 4, business: 4, said: 3, broadband: 3, continued: 2, networks: 2, industry: 2, echostar: 2, stock: 2, percent: 2

=== item_97 ===

said: 5, cascade: 4, company: 4, cut: 4, communications: 3, jobs: 3, corp: 2, statement: 2, percent: 2, workforce: 2

=== item_98 ===

said: 4, percent: 4, company: 3, software: 3, san: 2, rose: 2, year: 2, number: 2, billion: 2, imaging: 2

=== item_99 ===

company: 3, amazon: 3, stock: 3, selling: 2, pressure: 2, monday: 2, online: 2, companies: 2, facing: 2, investors: 2

Stylistic Analysis

Total unique words: 3080

Total word count: 17769

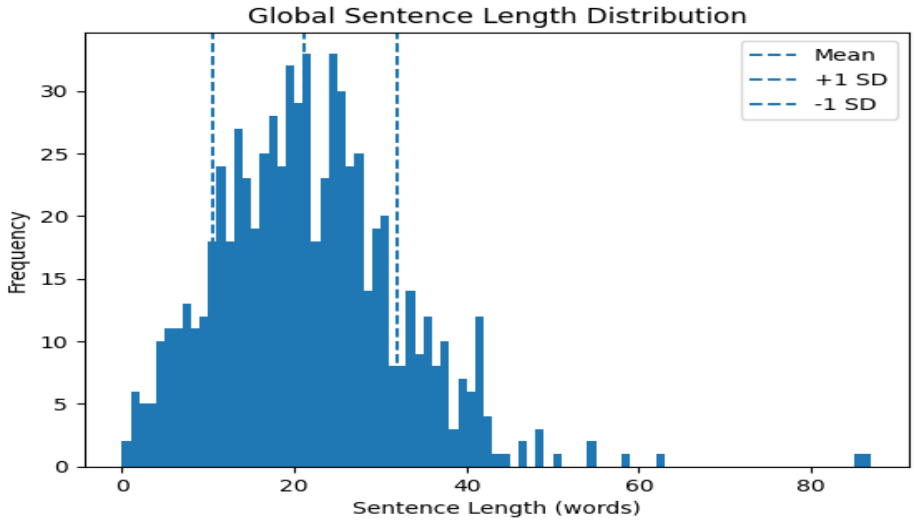
Type Token Ratio (TTR): 17.33%

Total stop-words count: 6774

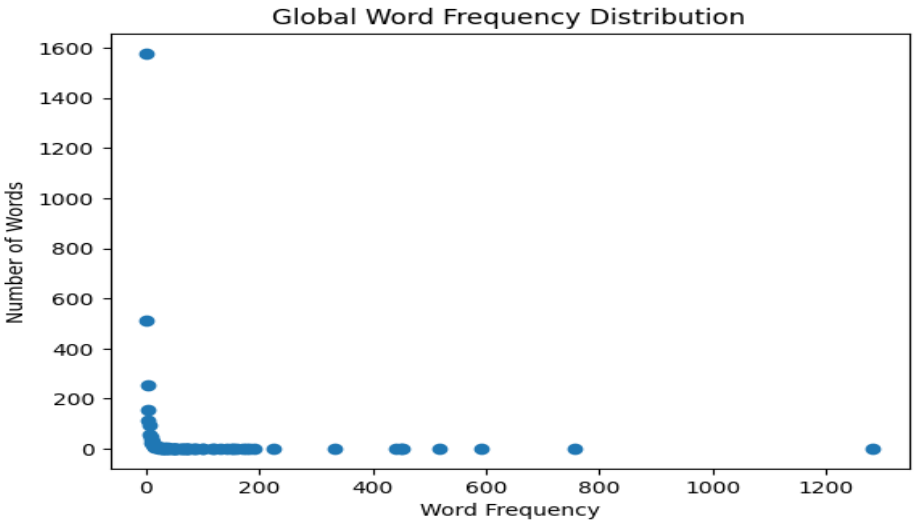
Most common Parts of Speech n-grams:

1. ('DT', 'NN'), 1146
2. ('NN', 'IN'), 939
3. ('JJ', 'NN'), 819
4. ('IN', 'DT'), 789
5. ('NN', 'NN'), 647
6. ('DT', 'JJ'), 537
7. ('NN', ','), 498
8. ('IN', 'NN'), 448
9. ('TO', 'VB'), 359
10. ('NN', '.'), 355
11. ('NN', 'VBD'), 259
12. ('.', 'DT'), 232
13. ('JJ', 'NNS'), 221
14. ('NN', 'VBZ'), 217
15. ('NNS', 'IN'), 208

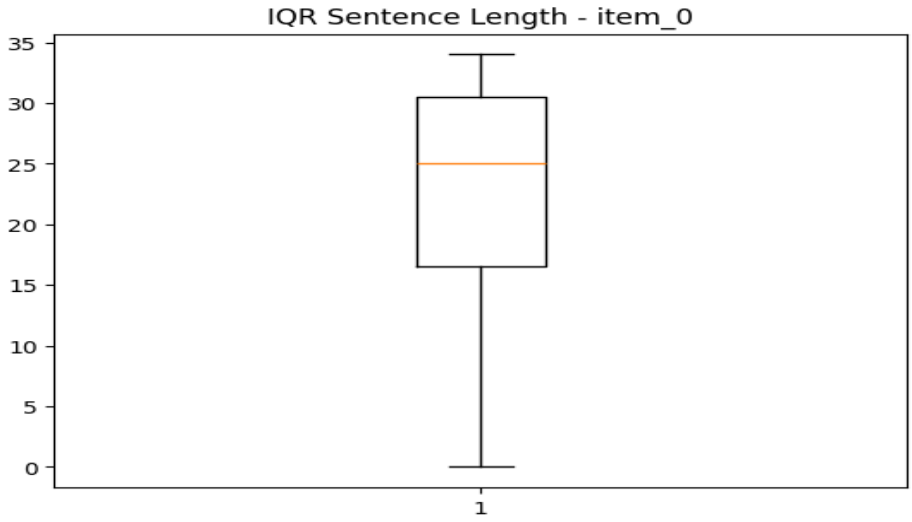
global_sentence_length_distribution.png



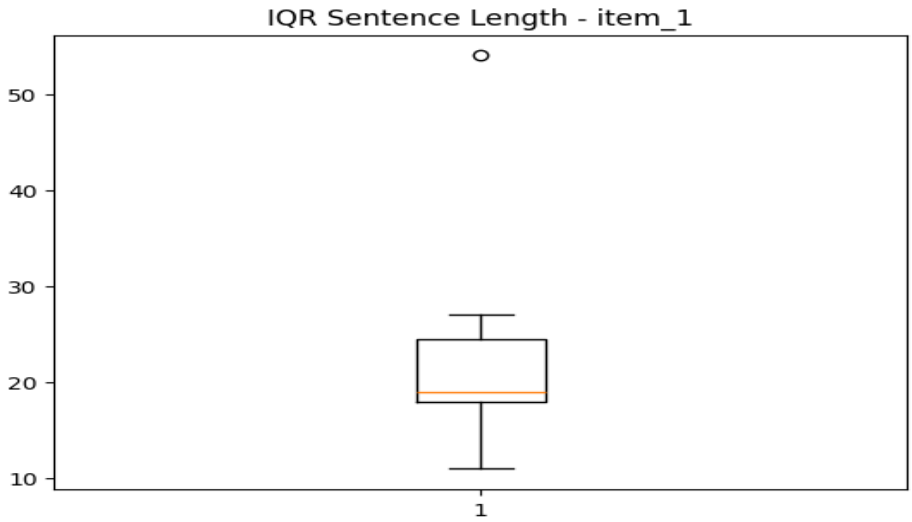
global_word_freq_distribution.png



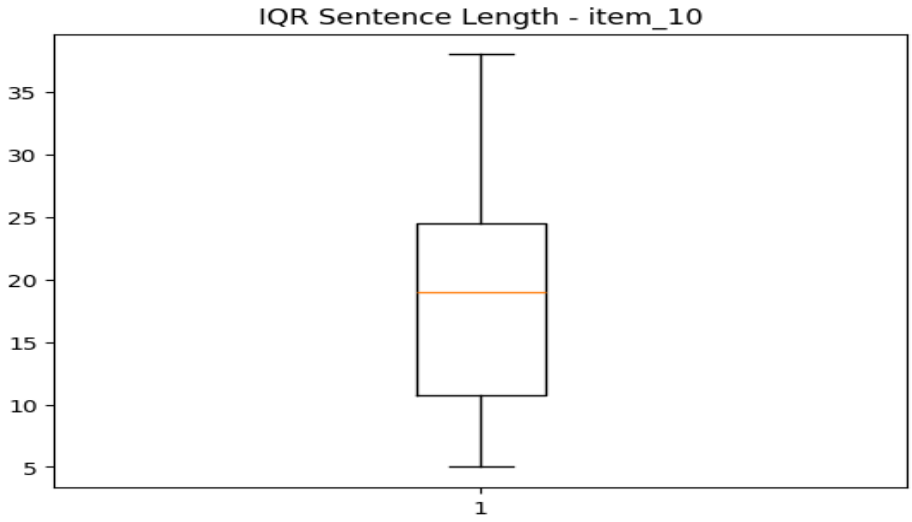
sentence_IQR_item_0.png



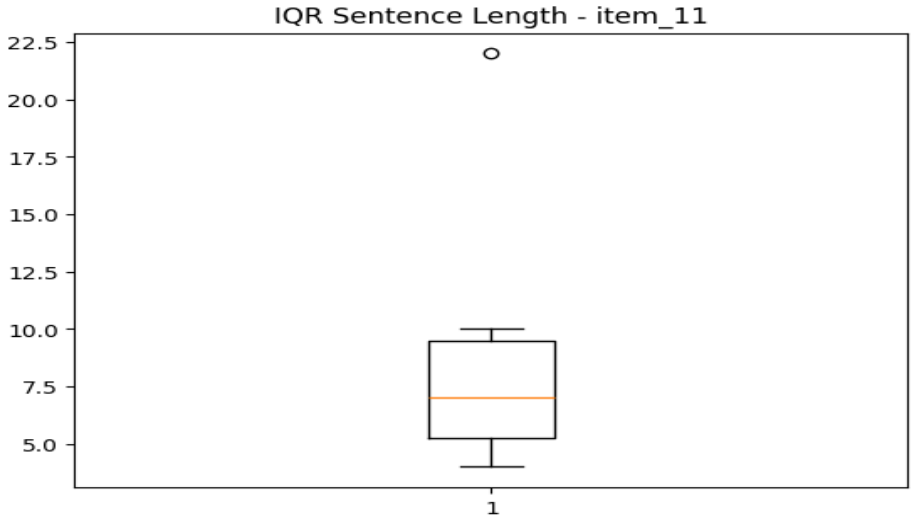
sentence_IQR_item_1.png



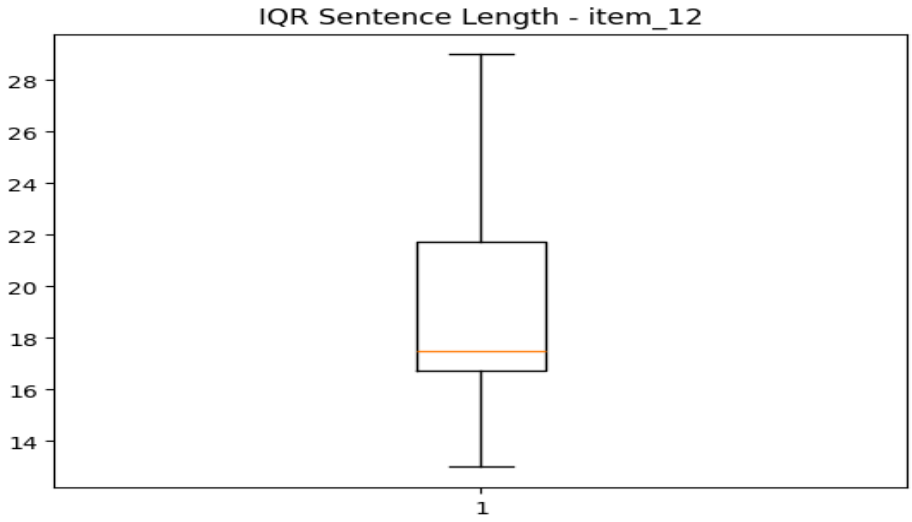
sentence_IQR_item_10.png



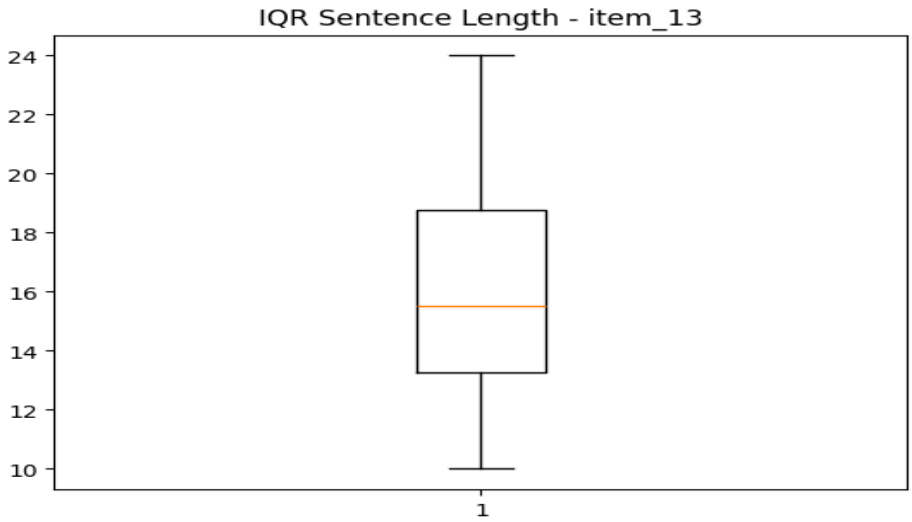
sentence_IQR_item_11.png



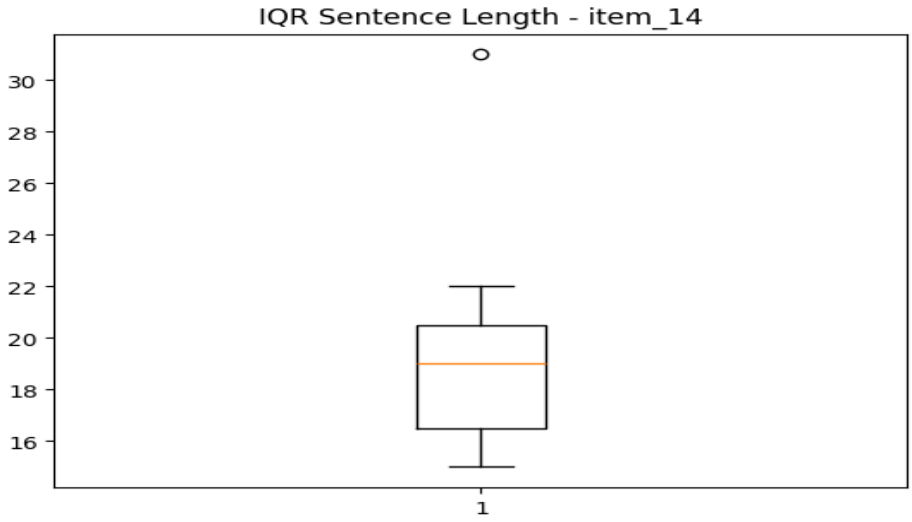
sentence_IQR_item_12.png



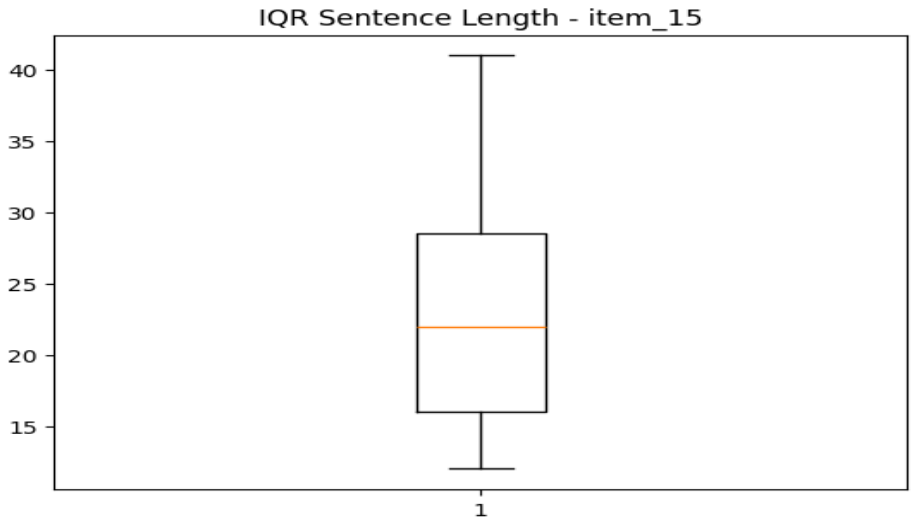
sentence_IQR_item_13.png



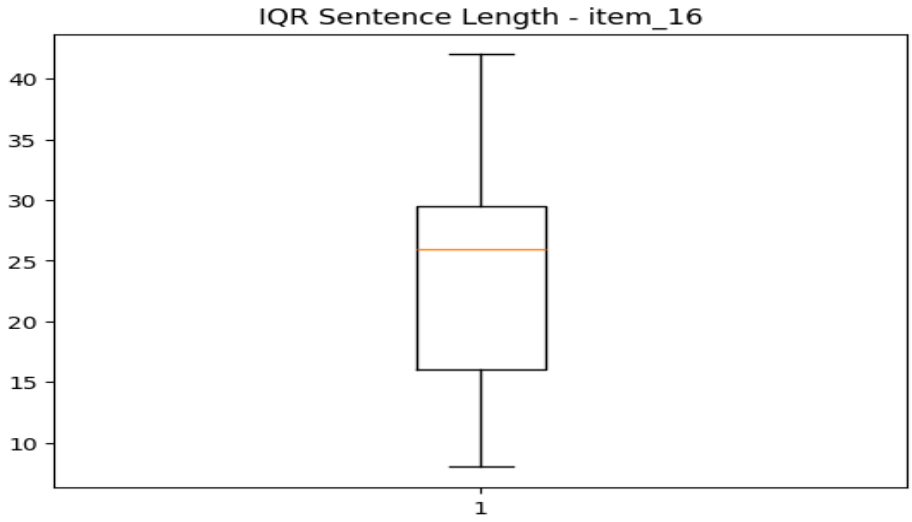
sentence_IQR_item_14.png



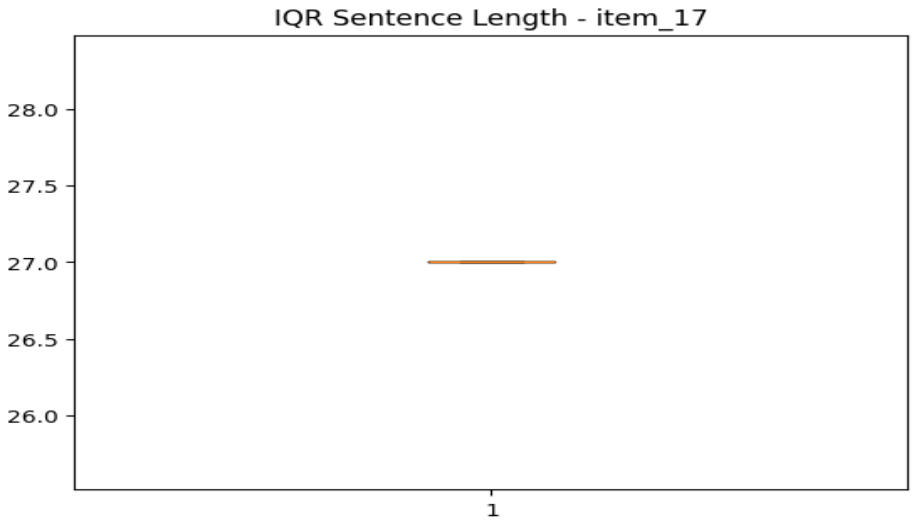
sentence_IQR_item_15.png



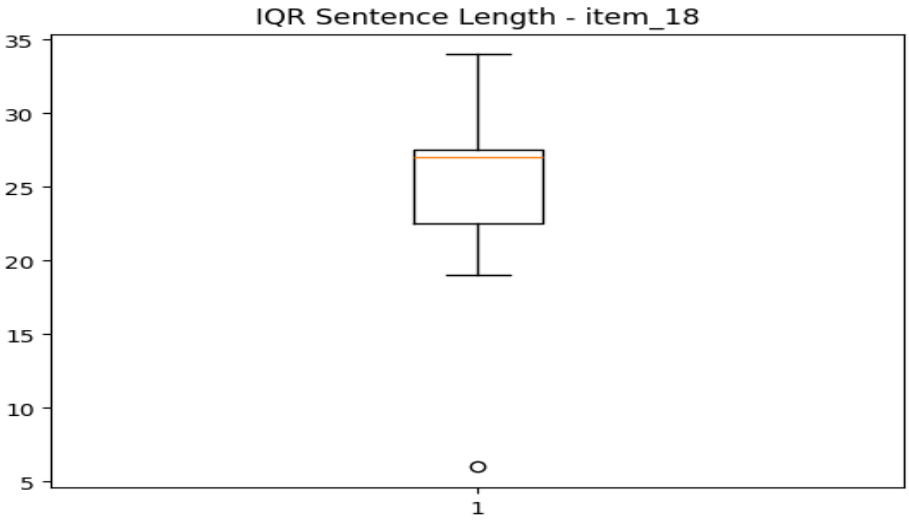
sentence_IQR_item_16.png



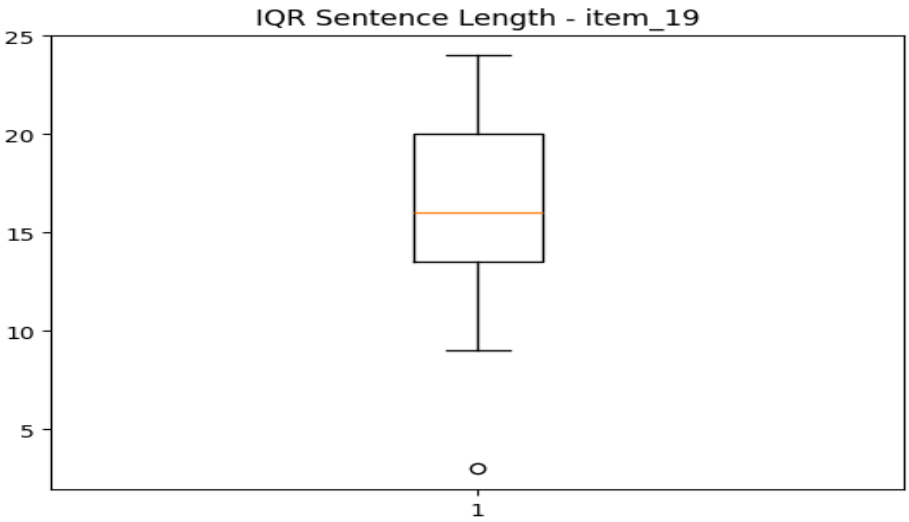
sentence_IQR_item_17.png



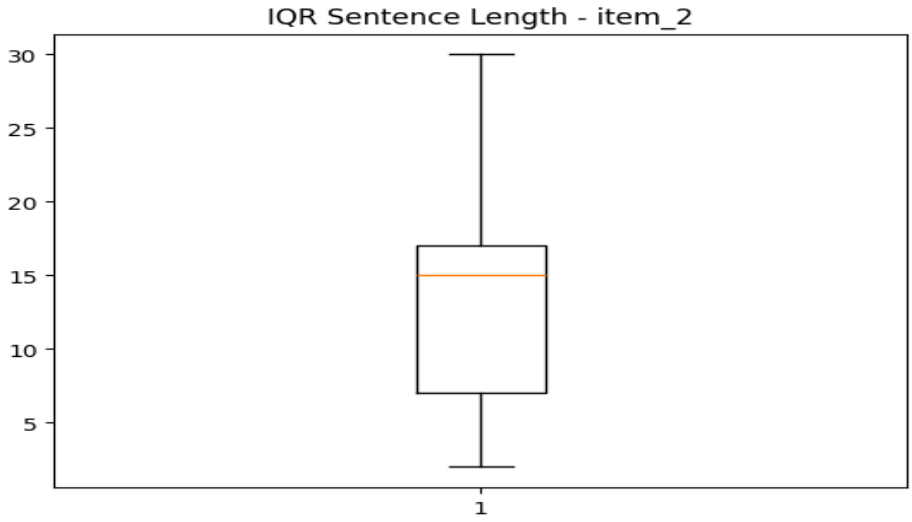
sentence_IQR_item_18.png



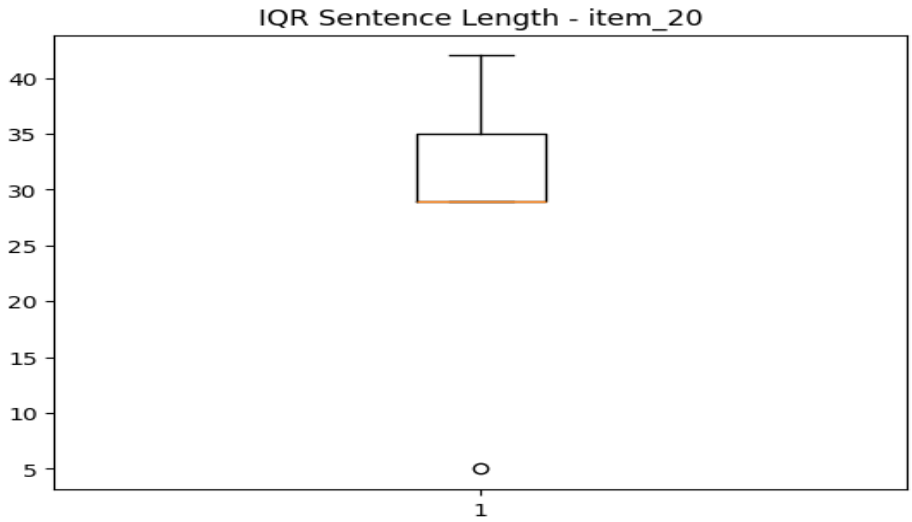
sentence_IQR_item_19.png



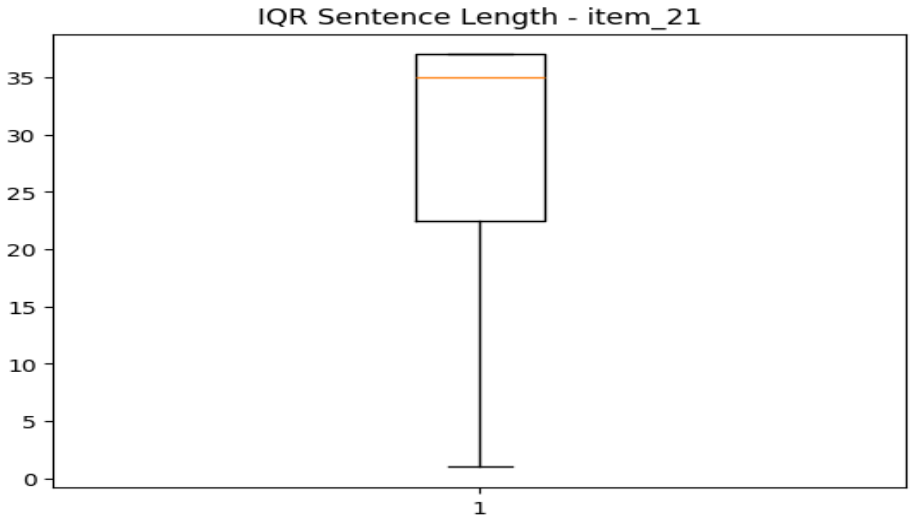
sentence_IQR_item_2.png



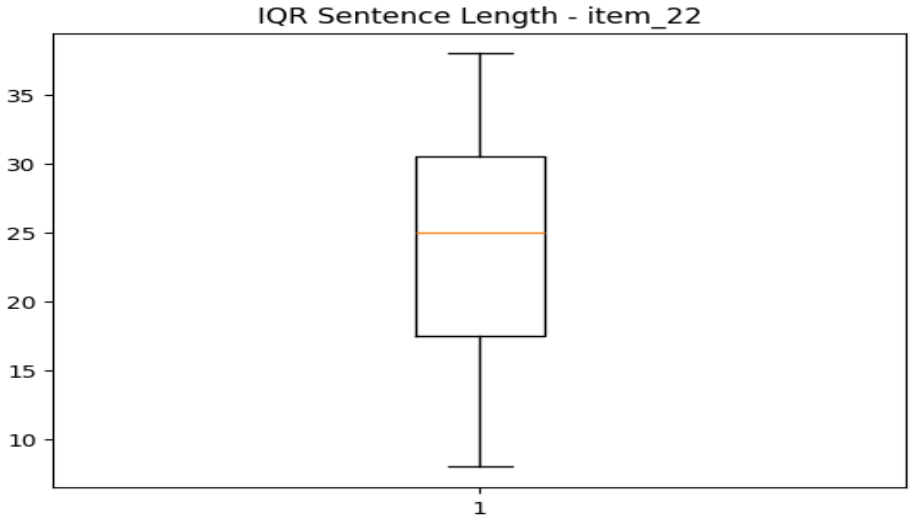
sentence_IQR_item_20.png



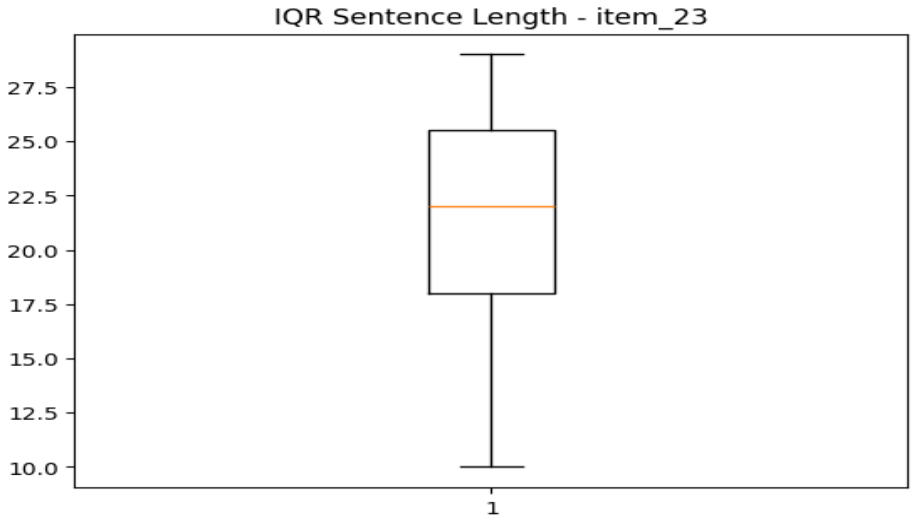
sentence_IQR_item_21.png



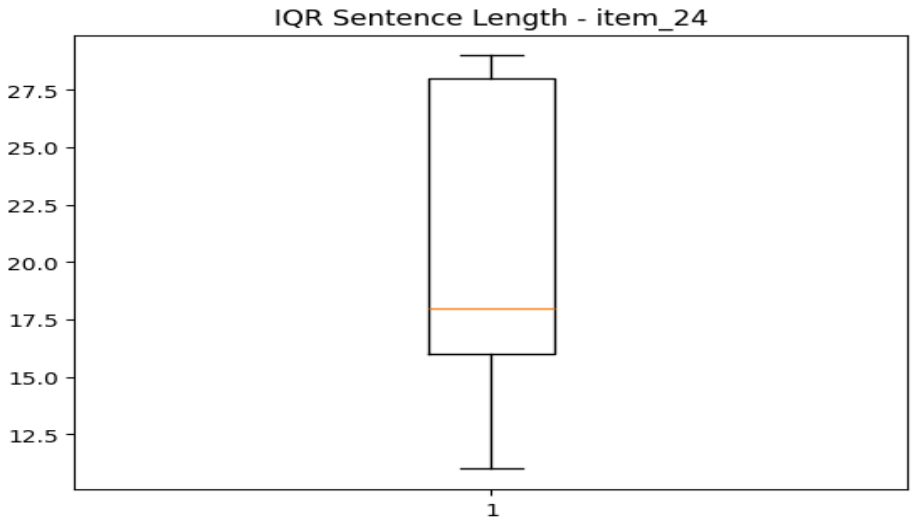
sentence_IQR_item_22.png



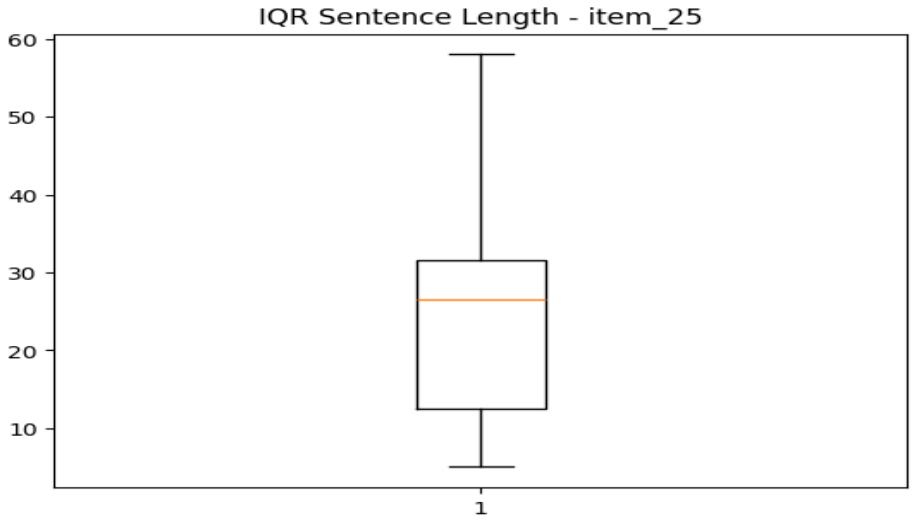
sentence_IQR_item_23.png



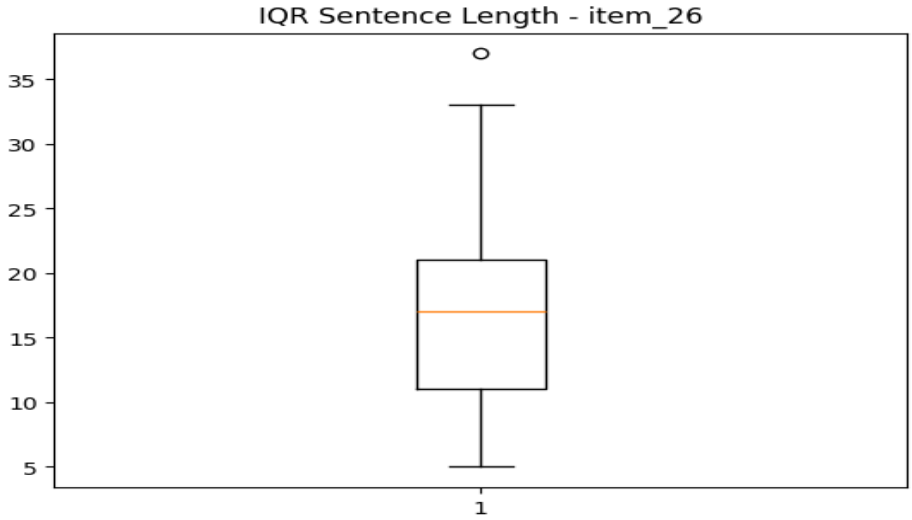
sentence_IQR_item_24.png



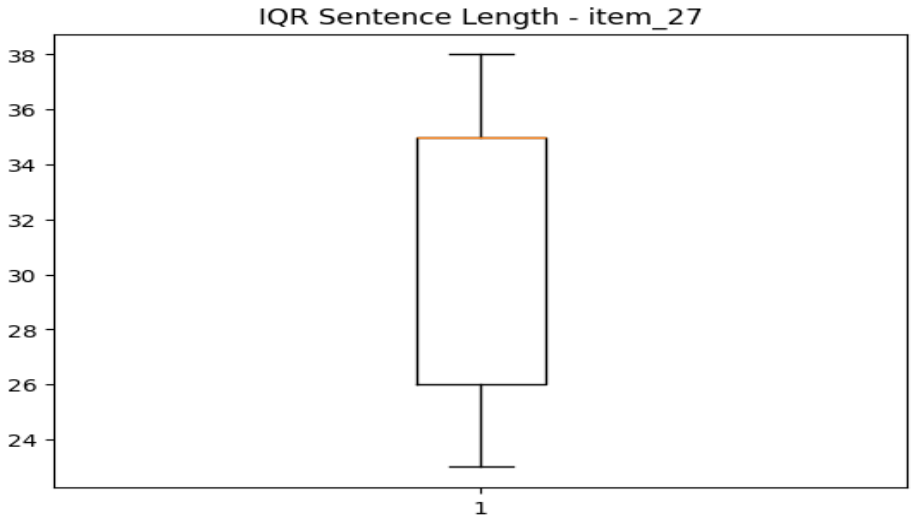
sentence_IQR_item_25.png



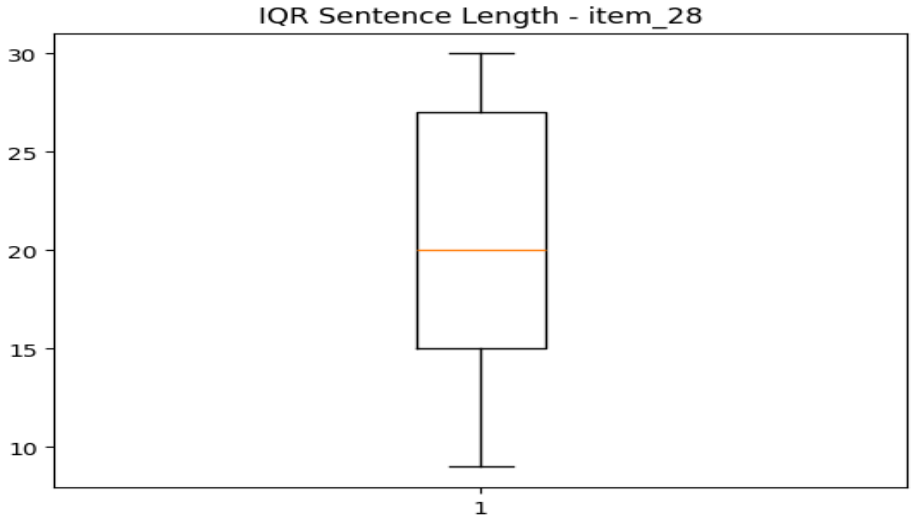
sentence_IQR_item_26.png



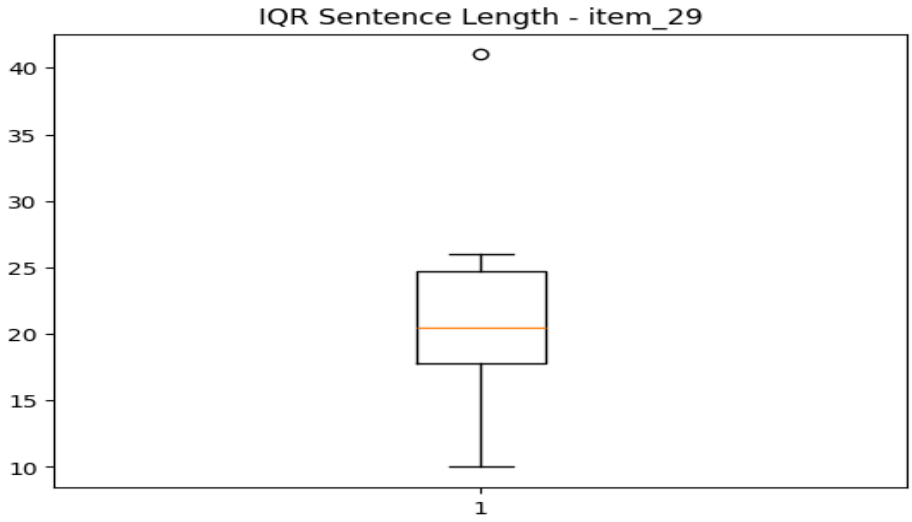
sentence_IQR_item_27.png



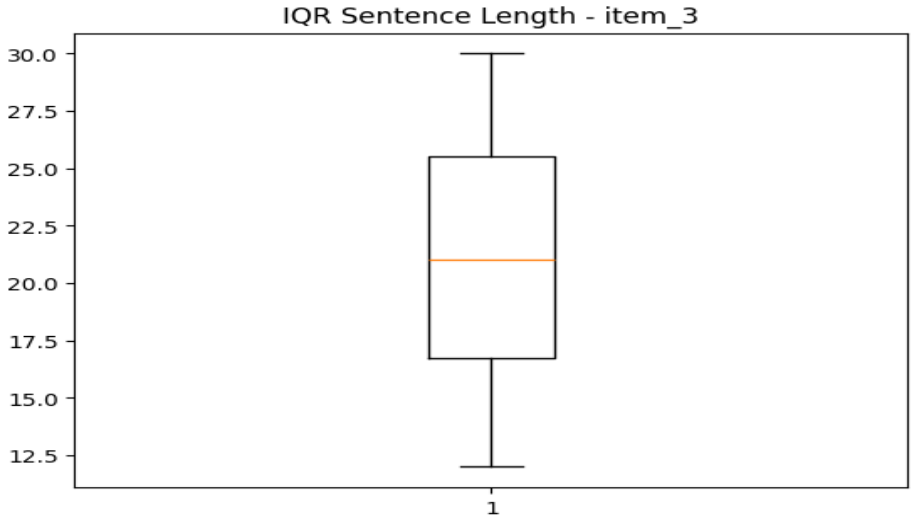
sentence_IQR_item_28.png



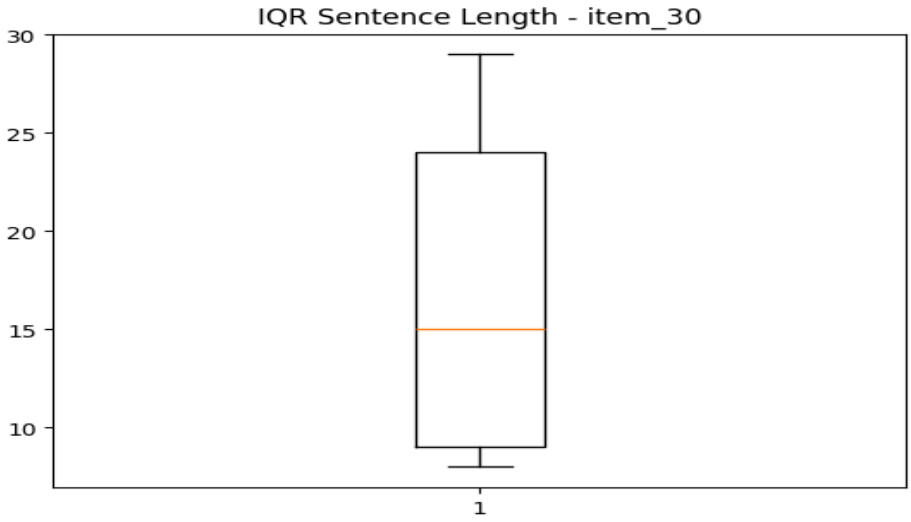
sentence_IQR_item_29.png



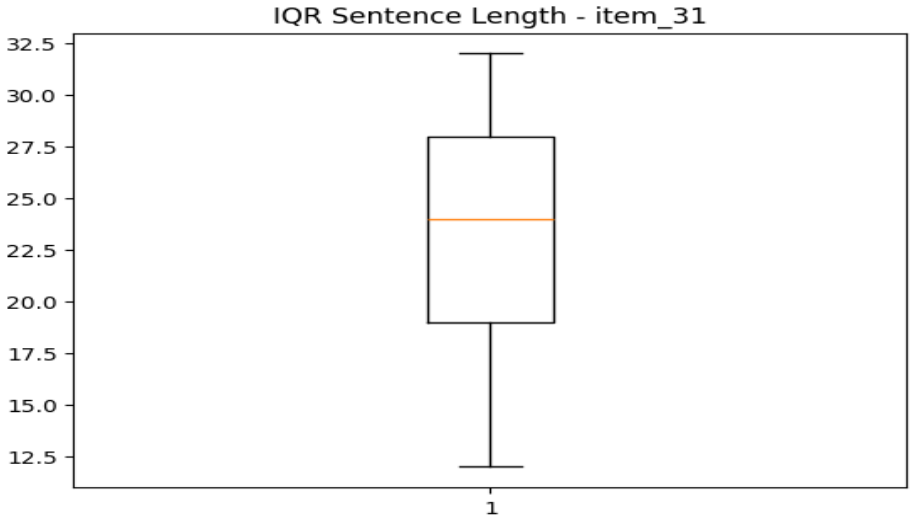
sentence_IQR_item_3.png



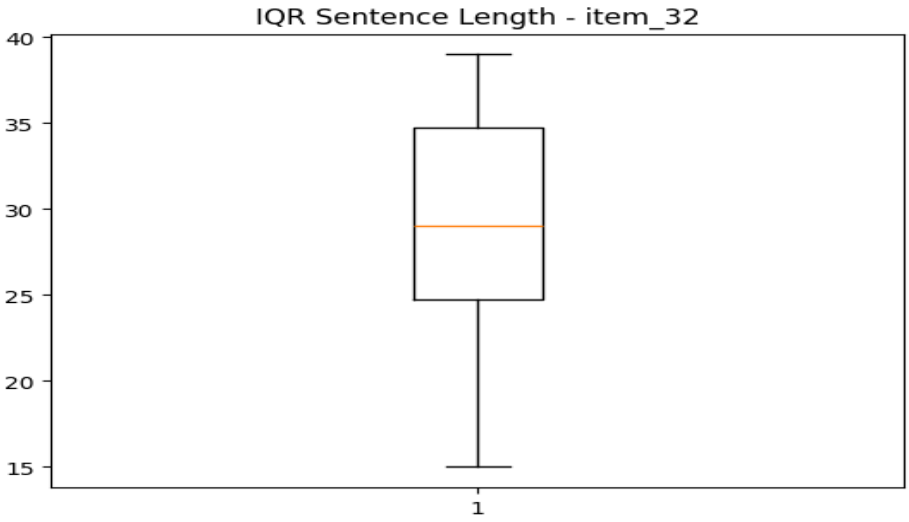
sentence_IQR_item_30.png



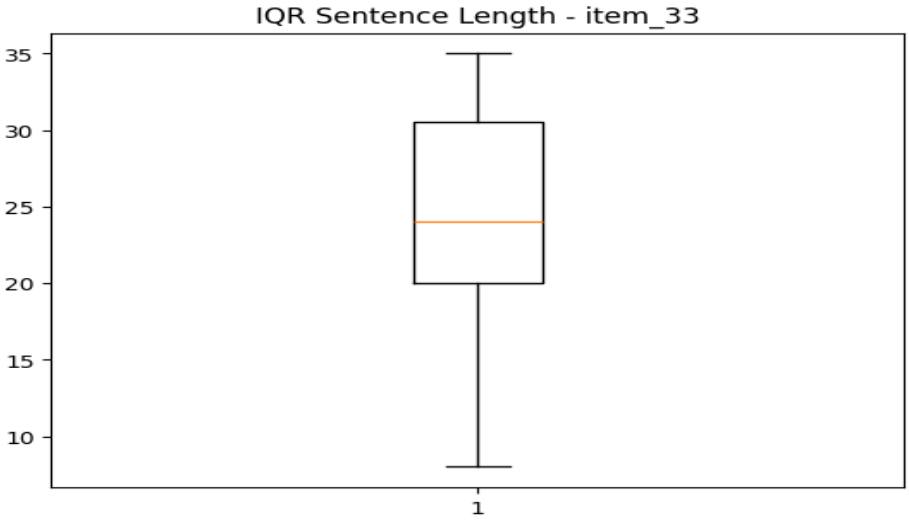
sentence_IQR_item_31.png



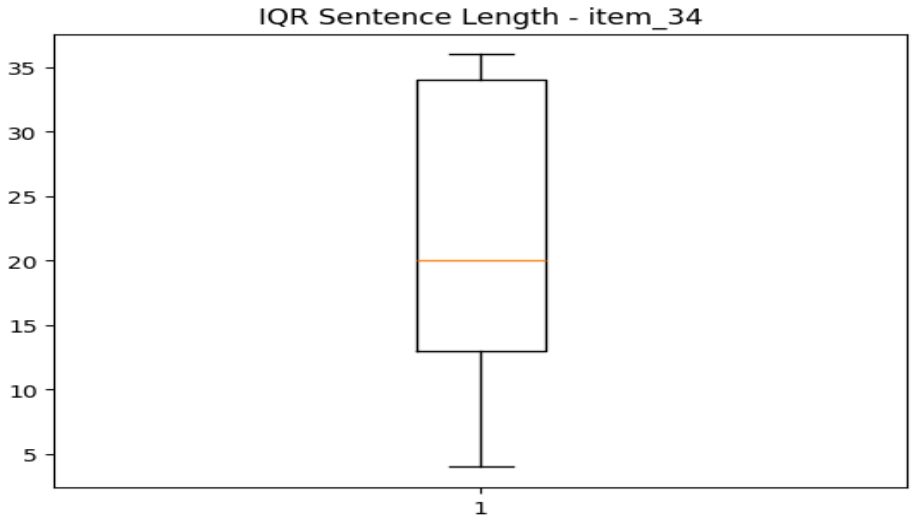
sentence_IQR_item_32.png



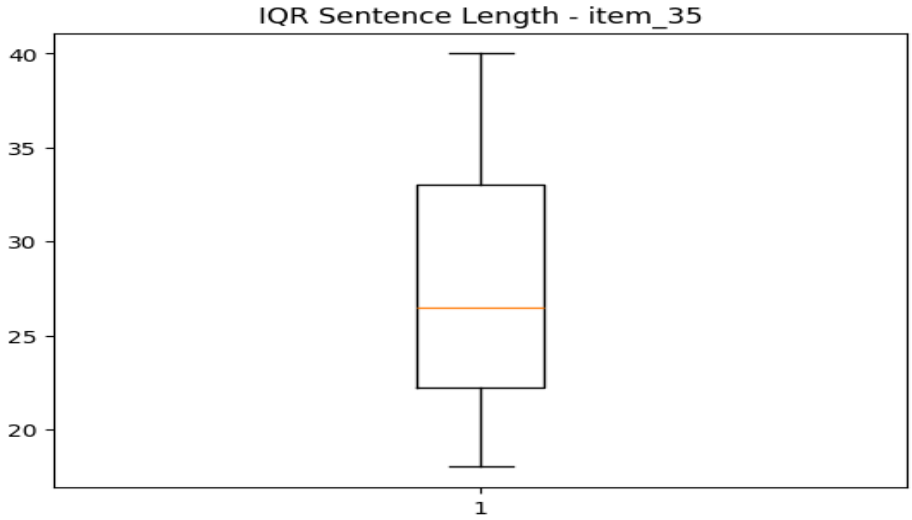
sentence_IQR_item_33.png



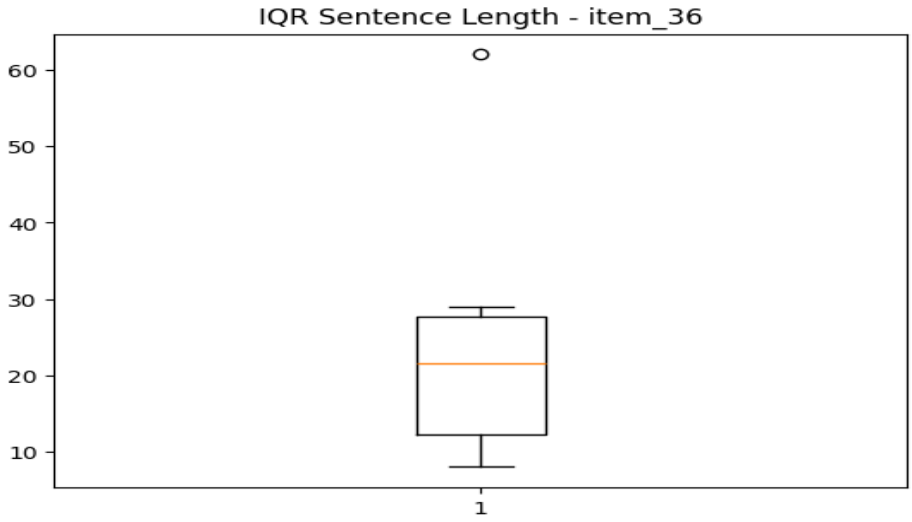
sentence_IQR_item_34.png



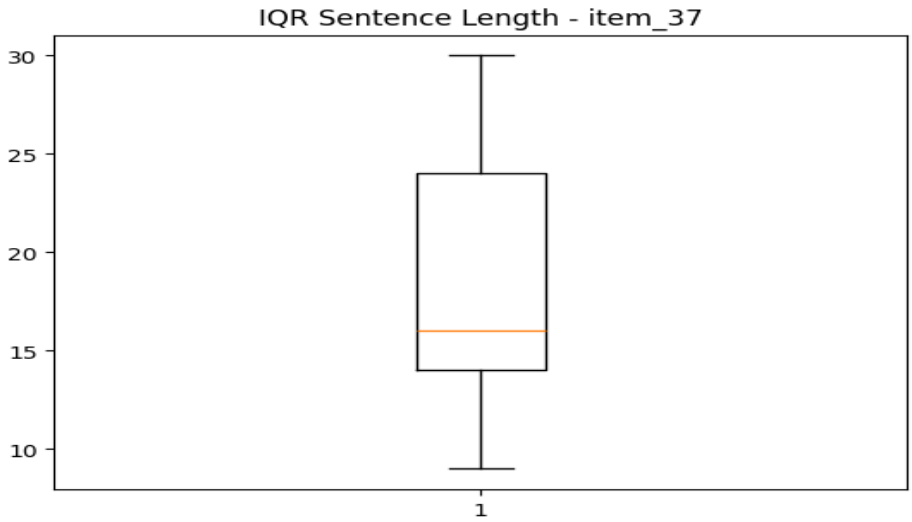
sentence_IQR_item_35.png



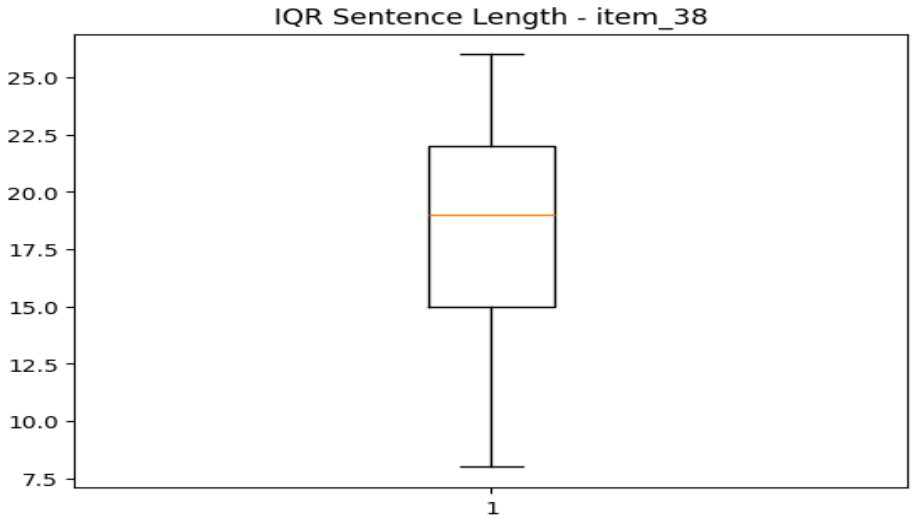
sentence_IQR_item_36.png



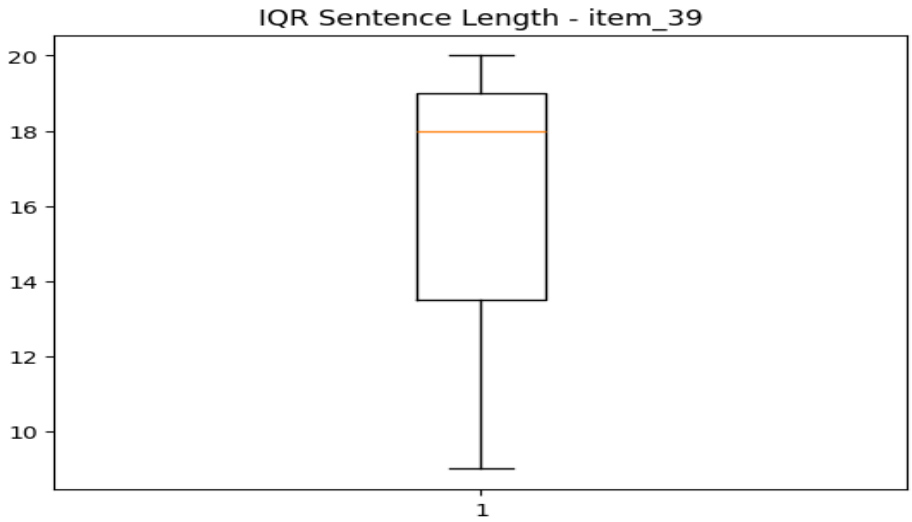
sentence_IQR_item_37.png



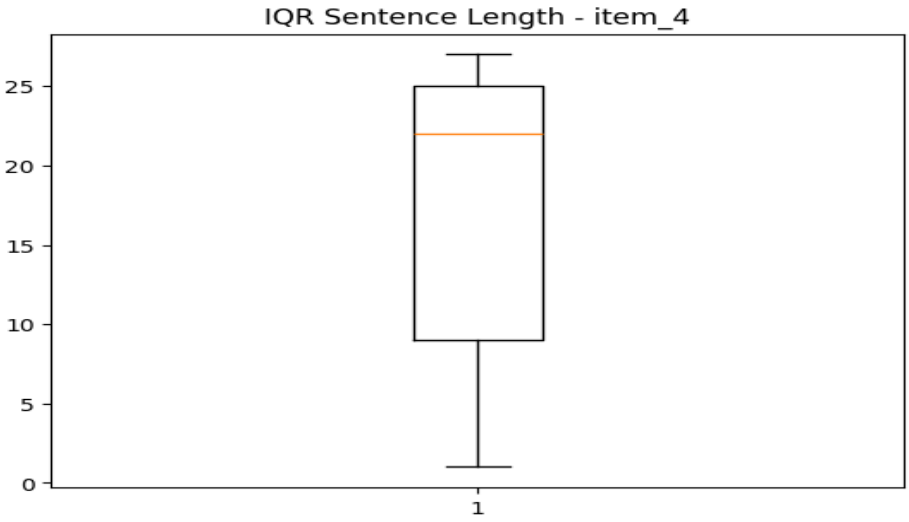
sentence_IQR_item_38.png



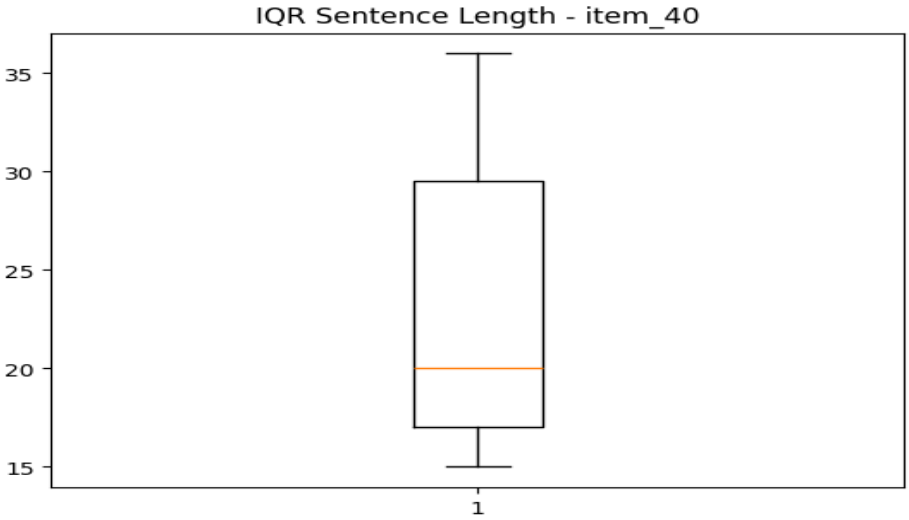
sentence_IQR_item_39.png



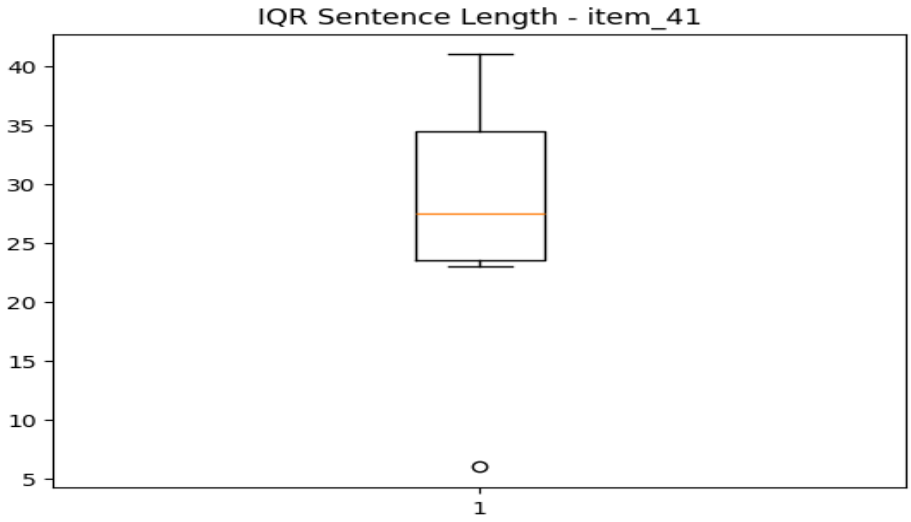
sentence_IQR_item_4.png



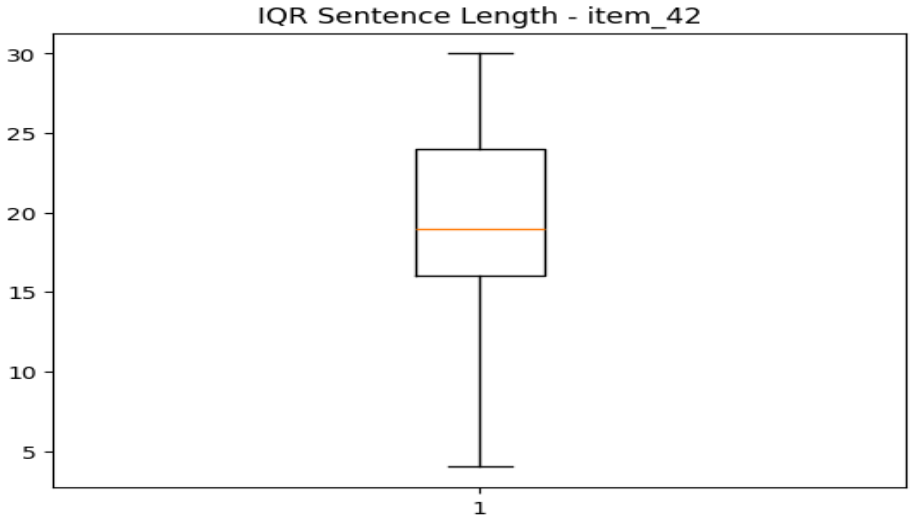
sentence_IQR_item_40.png



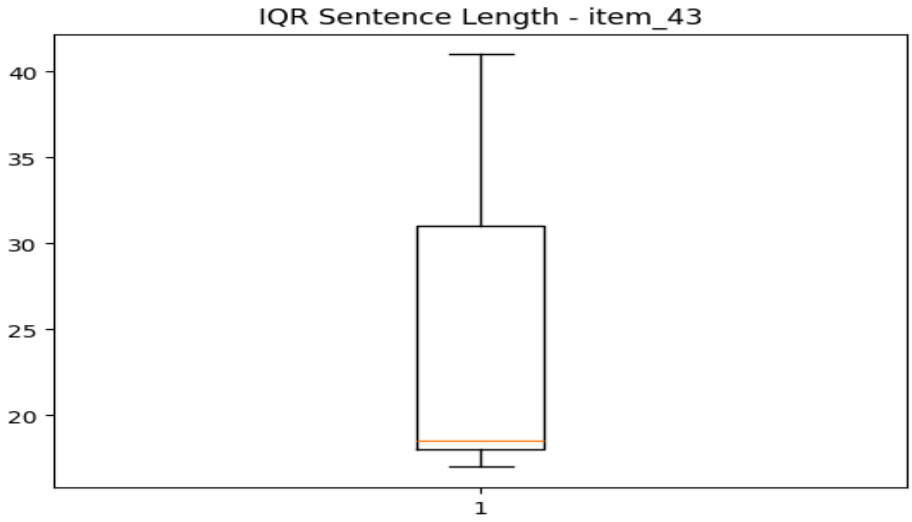
sentence_IQR_item_41.png



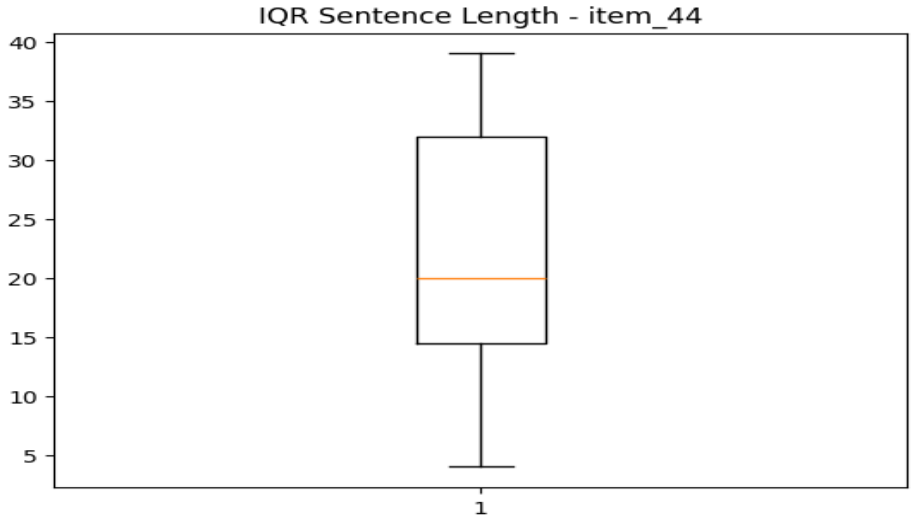
sentence_IQR_item_42.png



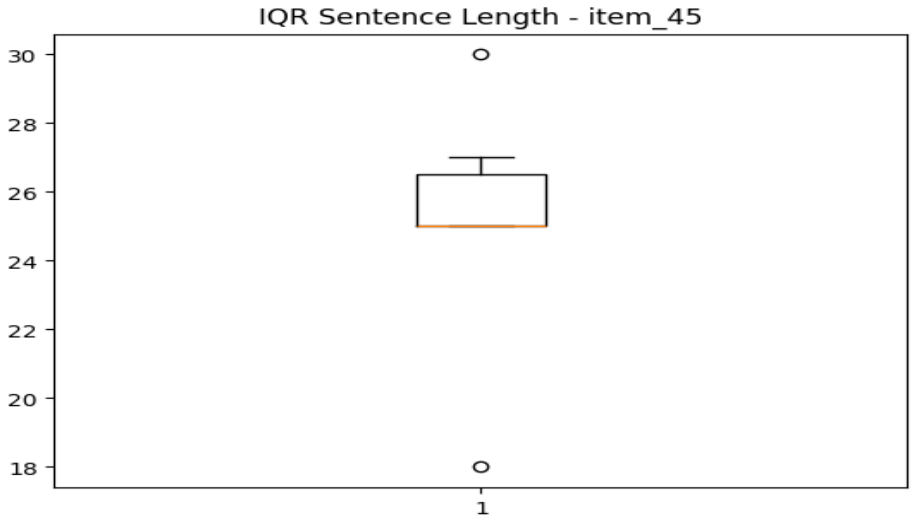
sentence_IQR_item_43.png



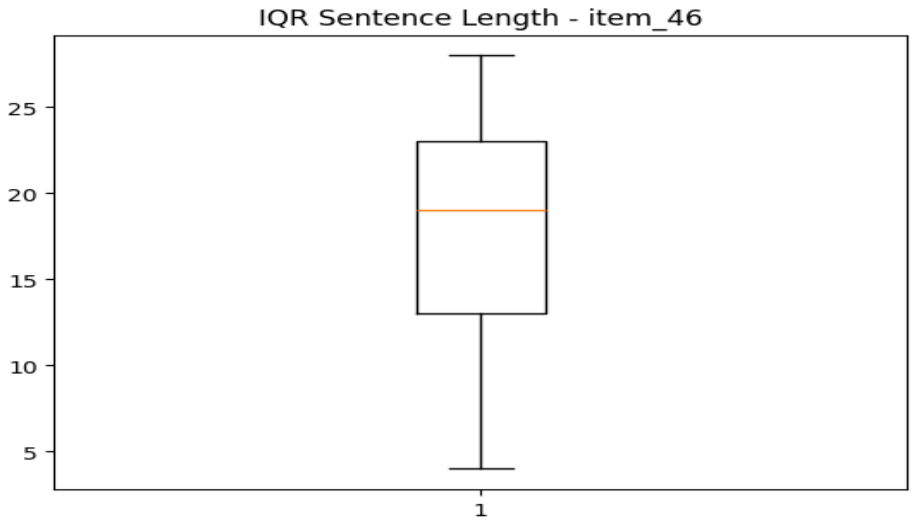
sentence_IQR_item_44.png



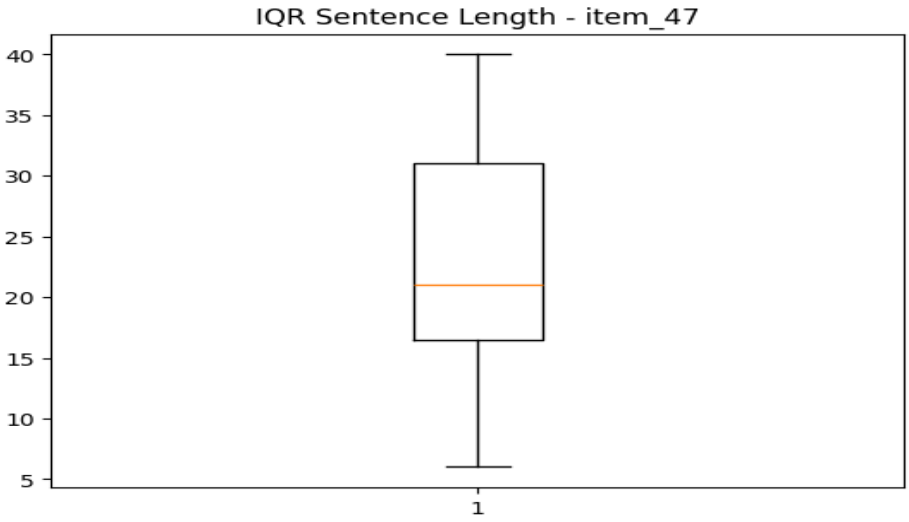
sentence_IQR_item_45.png



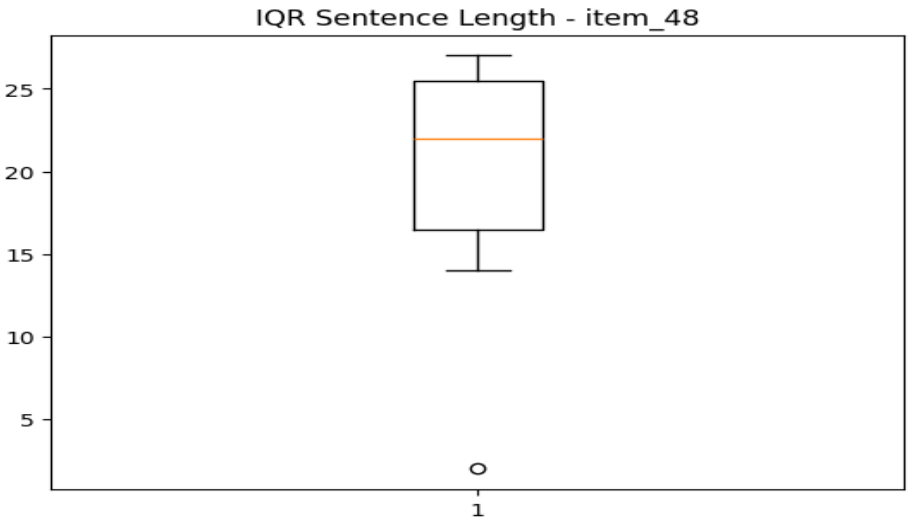
sentence_IQR_item_46.png



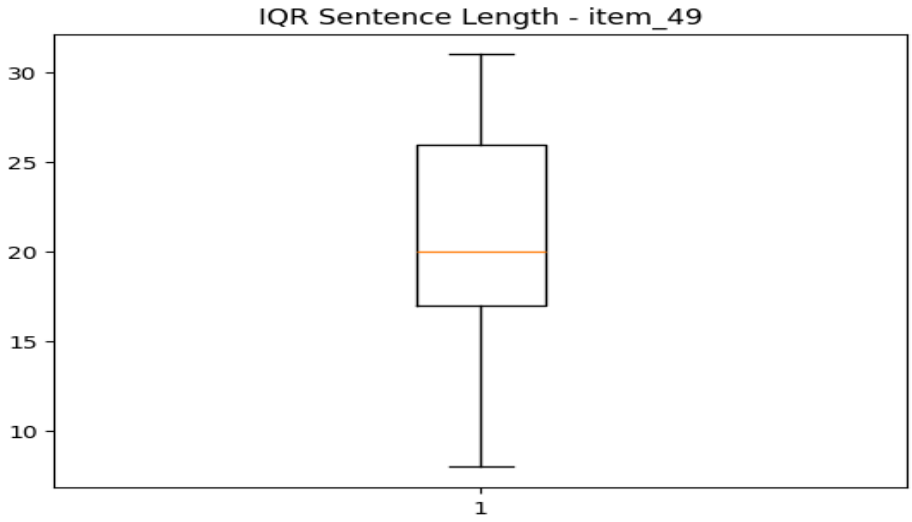
sentence_IQR_item_47.png



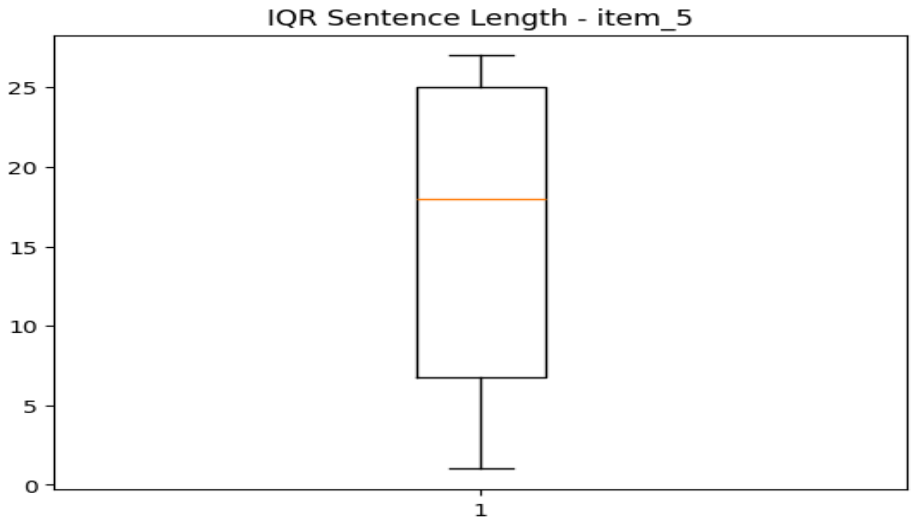
sentence_IQR_item_48.png



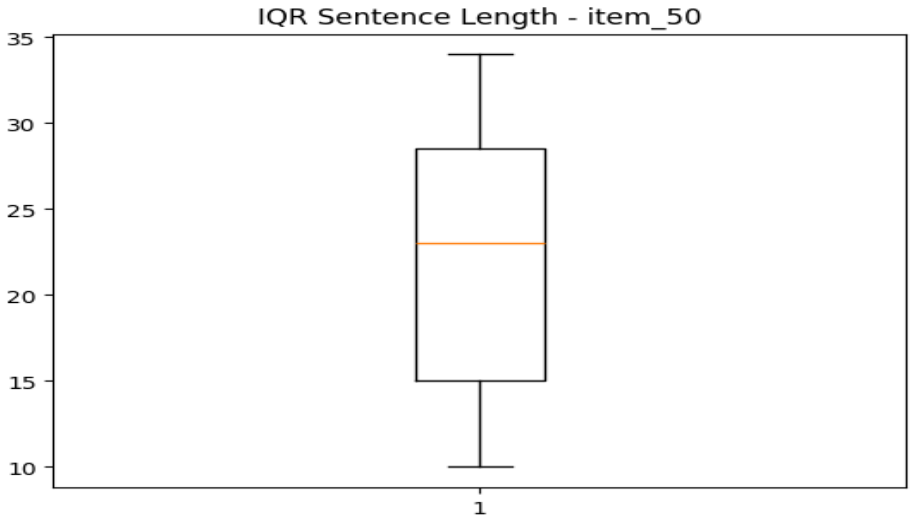
sentence_IQR_item_49.png



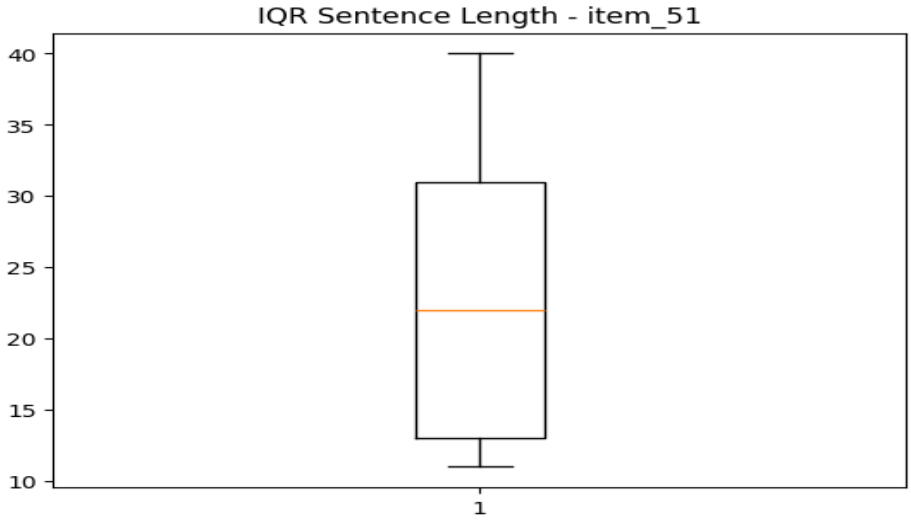
sentence_IQR_item_5.png



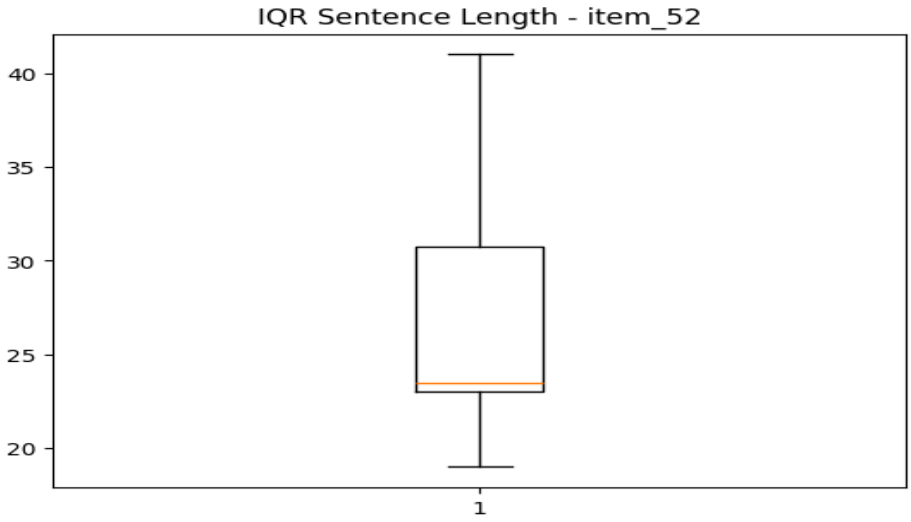
sentence_IQR_item_50.png



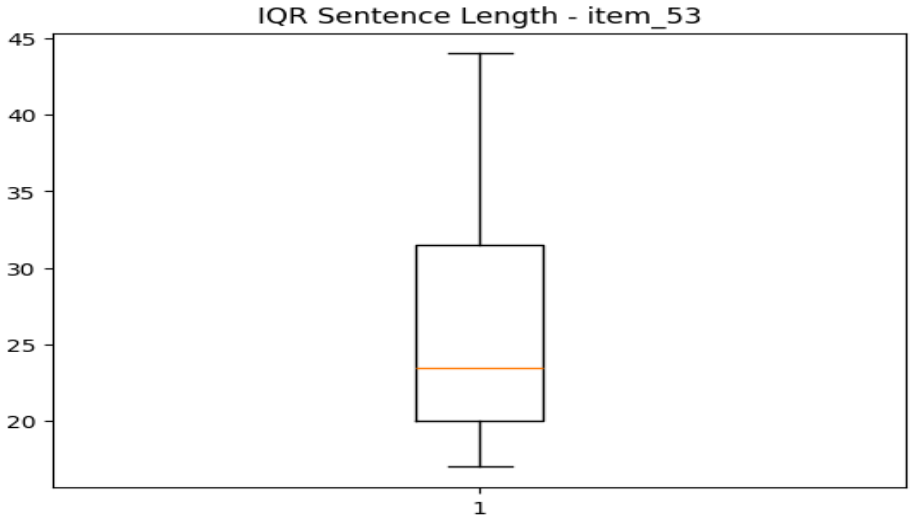
sentence_IQR_item_51.png



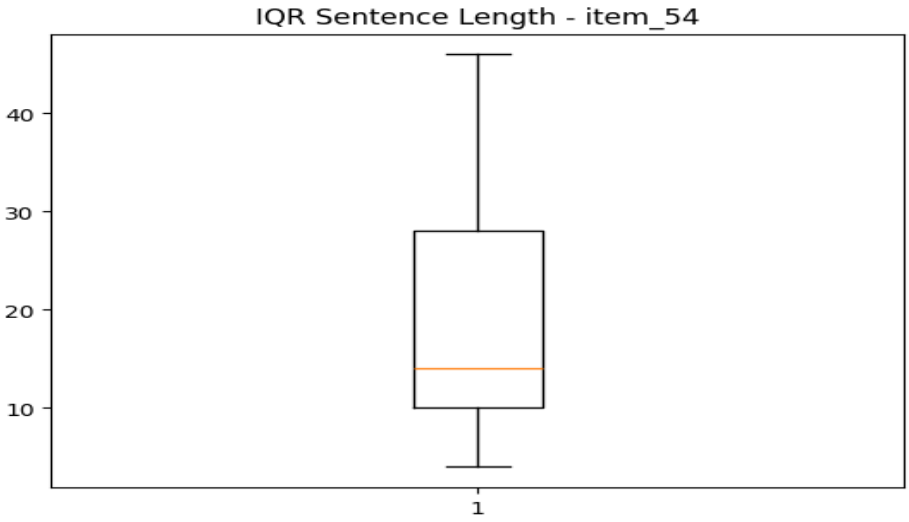
sentence_IQR_item_52.png



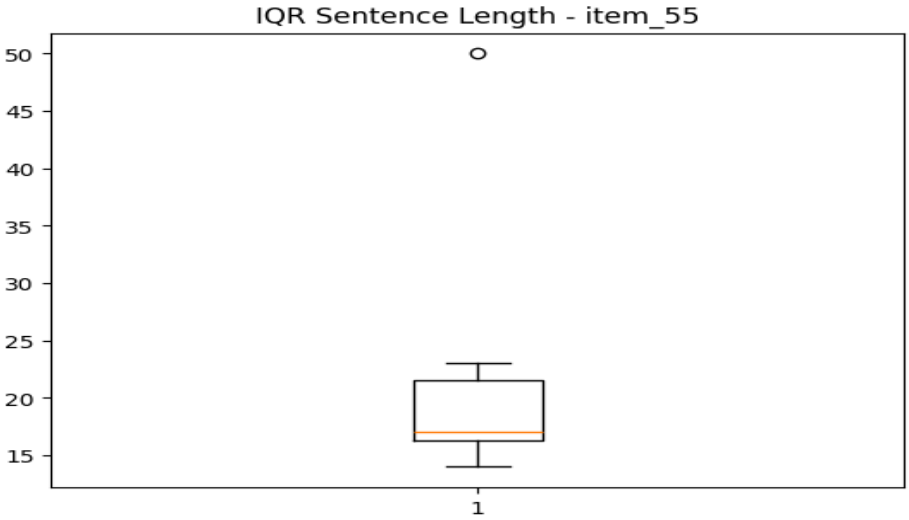
sentence_IQR_item_53.png



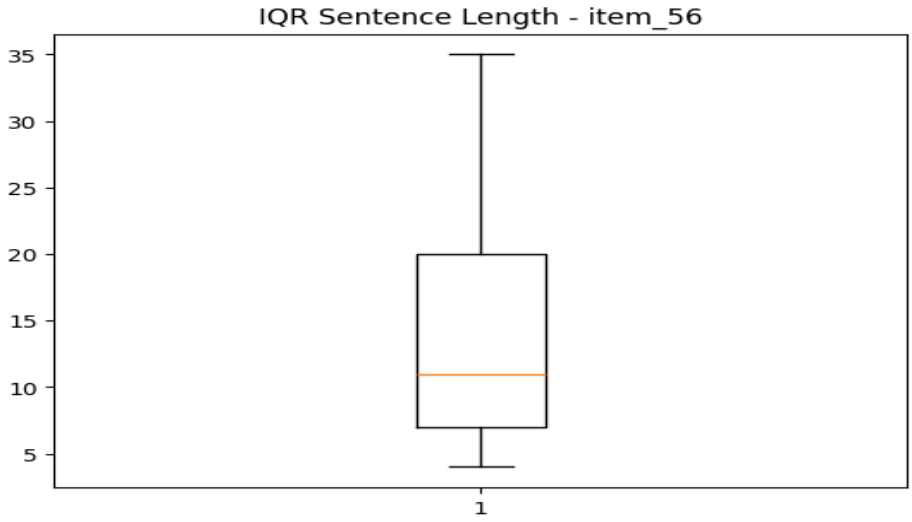
sentence_IQR_item_54.png



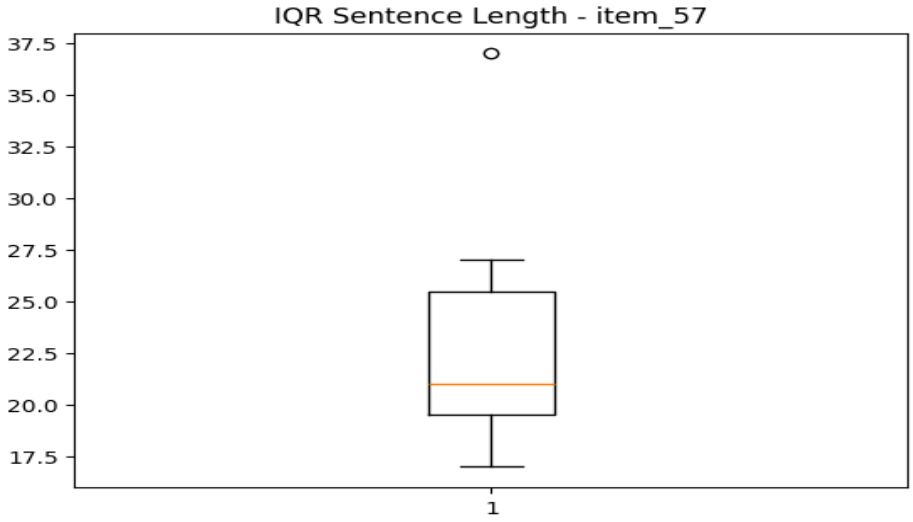
sentence_IQR_item_55.png



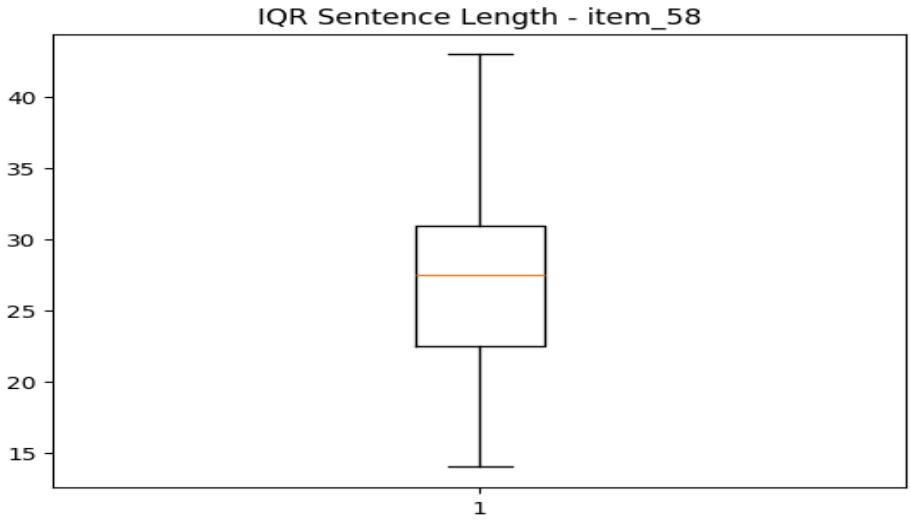
sentence_IQR_item_56.png



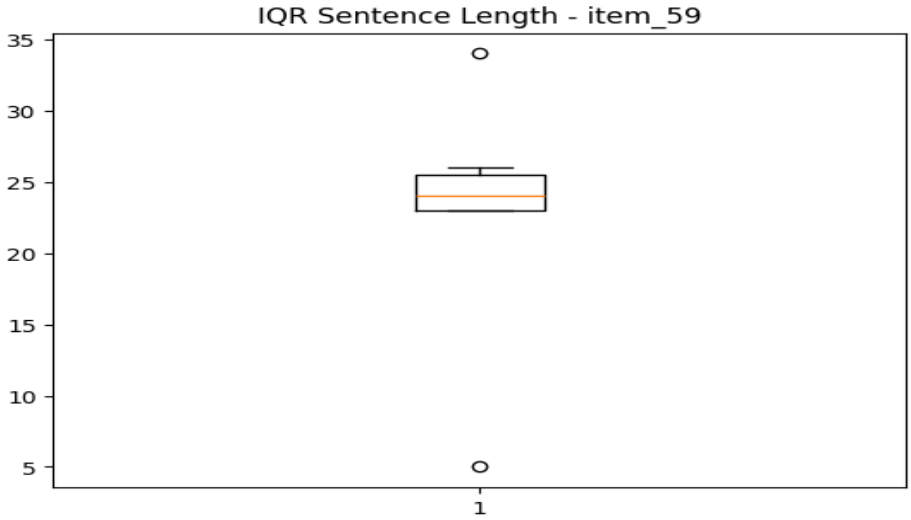
sentence_IQR_item_57.png



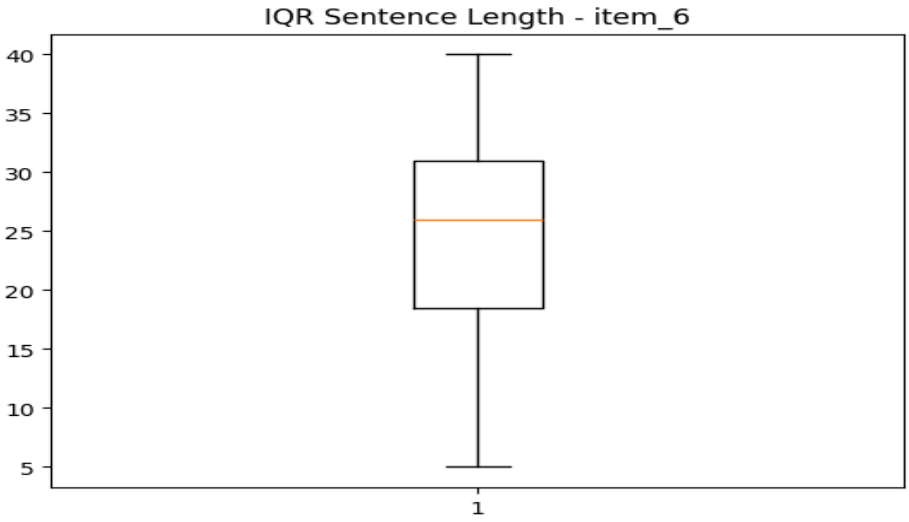
sentence_IQR_item_58.png



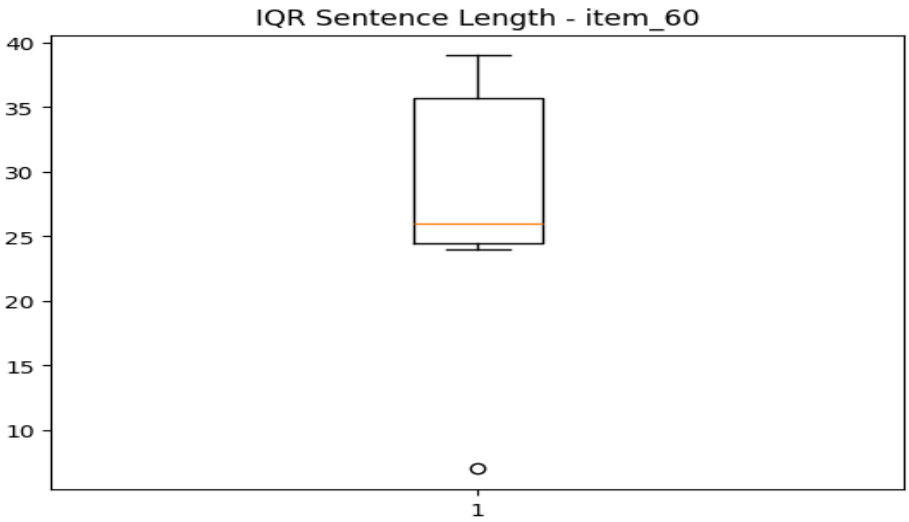
sentence_IQR_item_59.png



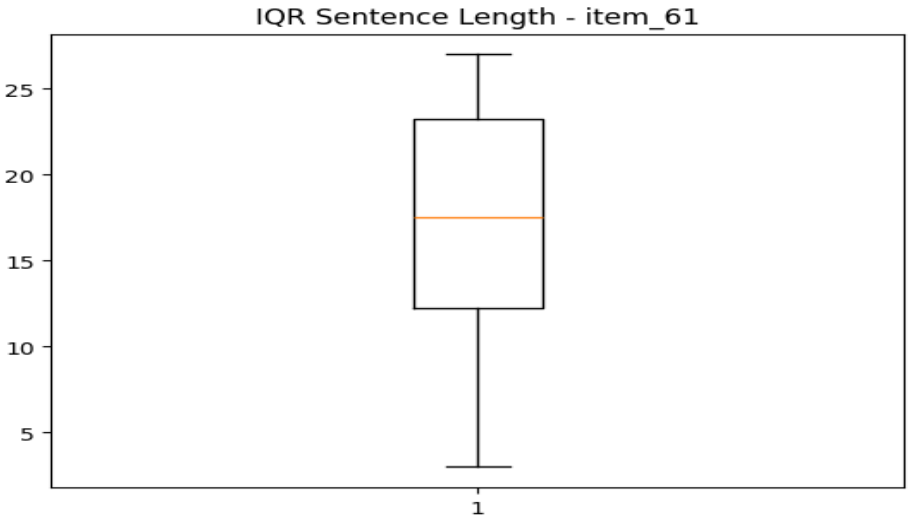
sentence_IQR_item_6.png



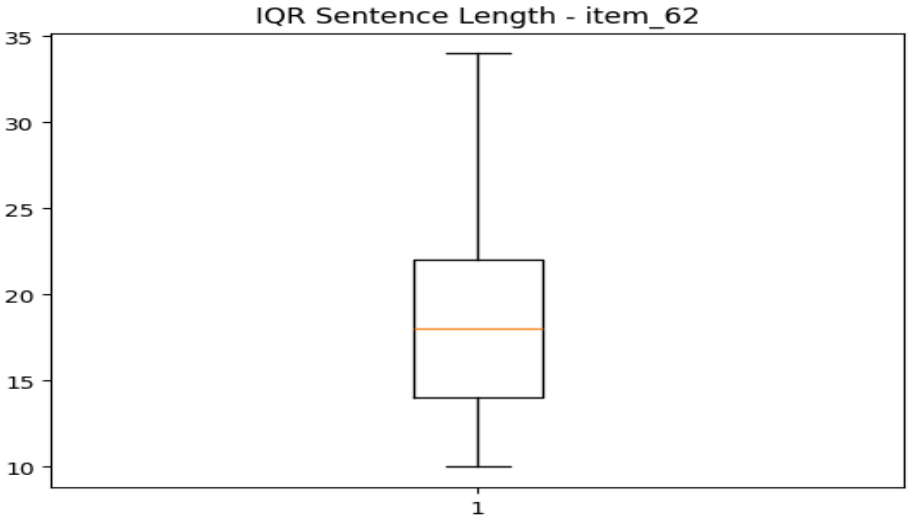
sentence_IQR_item_60.png



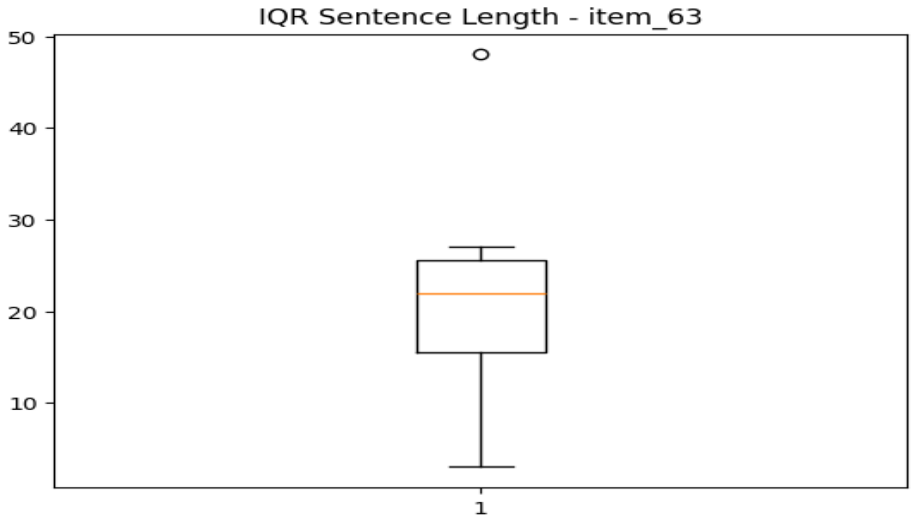
sentence_IQR_item_61.png



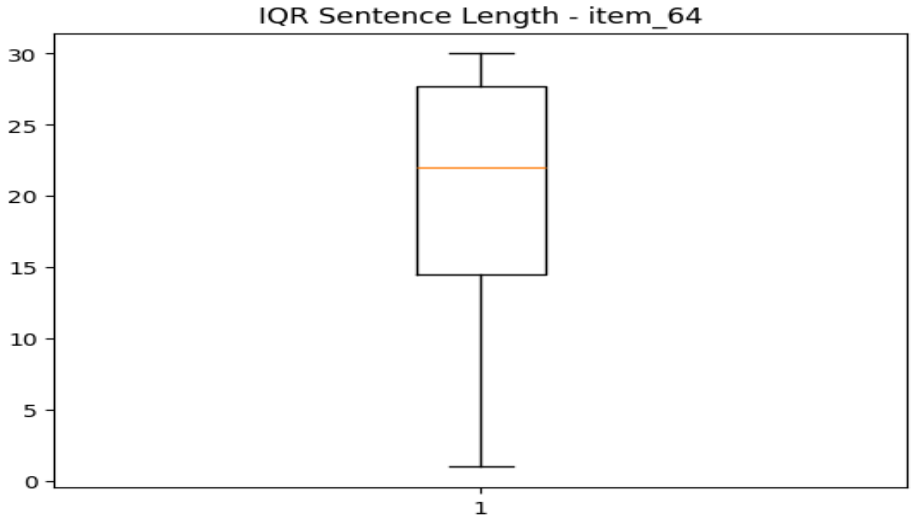
sentence_IQR_item_62.png



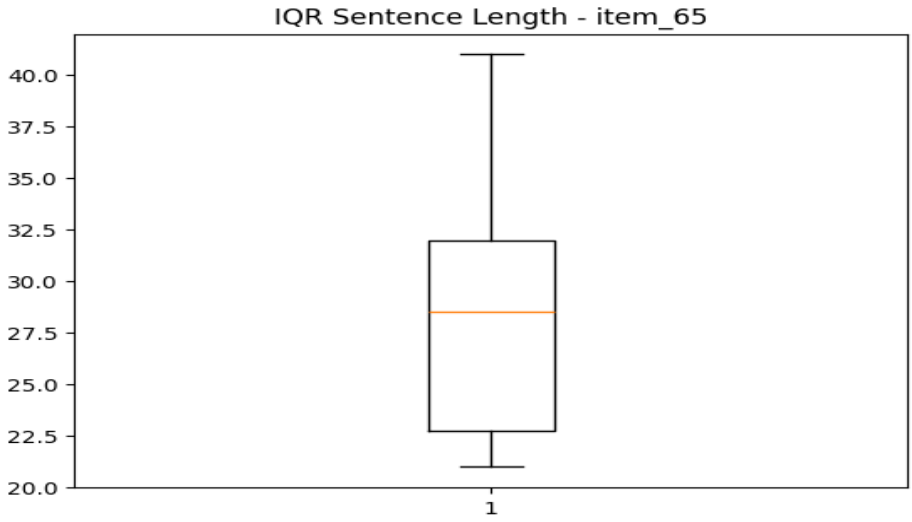
sentence_IQR_item_63.png



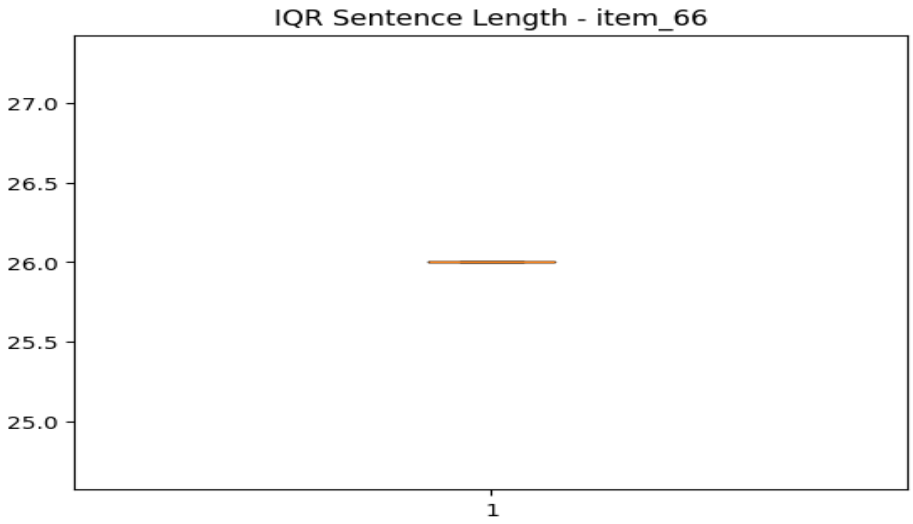
sentence_IQR_item_64.png



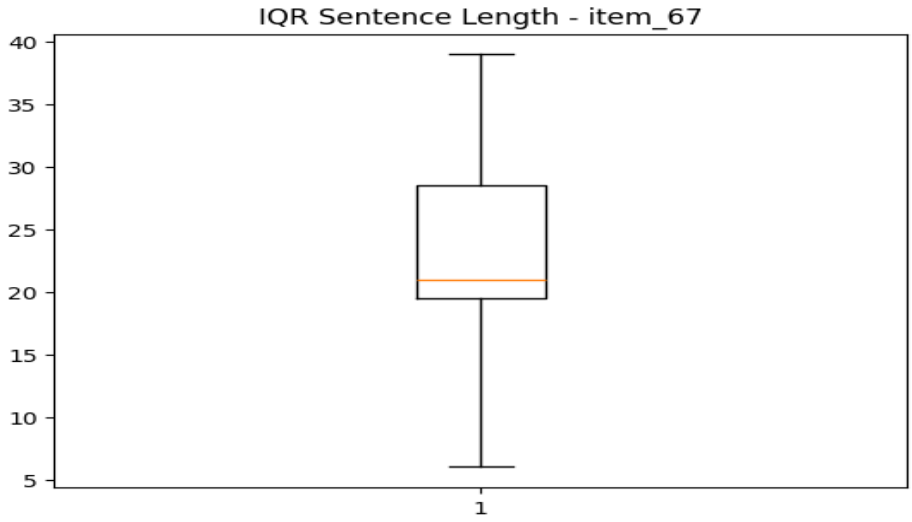
sentence_IQR_item_65.png



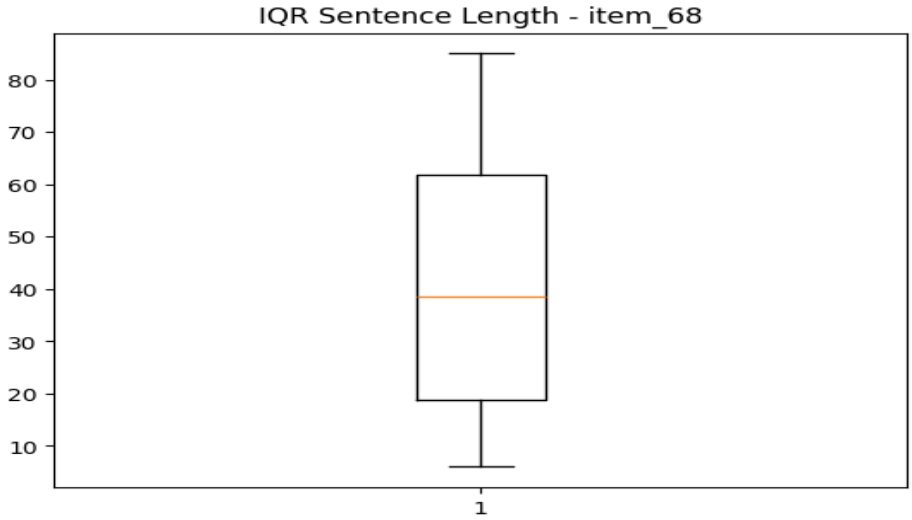
sentence_IQR_item_66.png



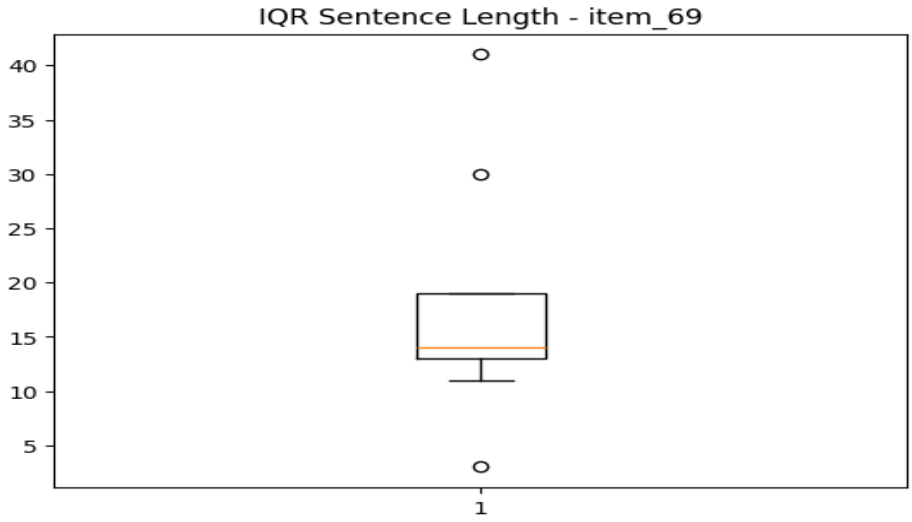
sentence_IQR_item_67.png



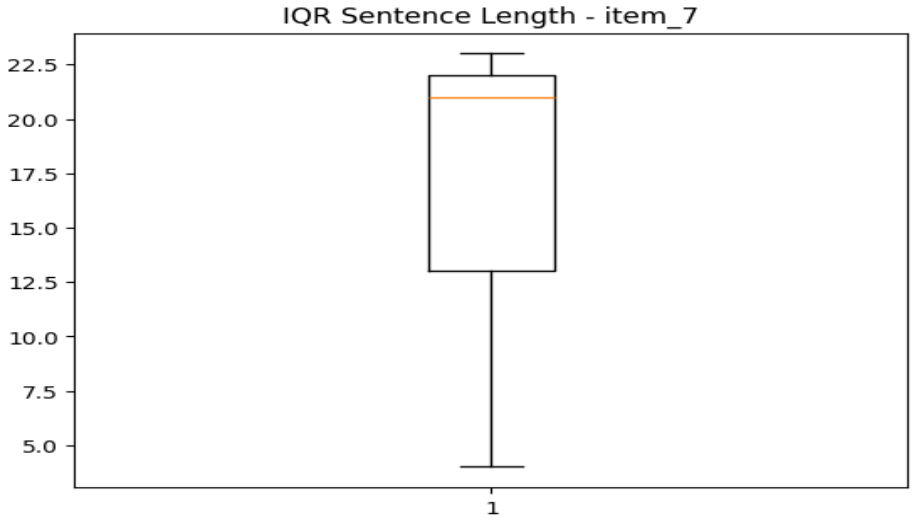
sentence_IQR_item_68.png



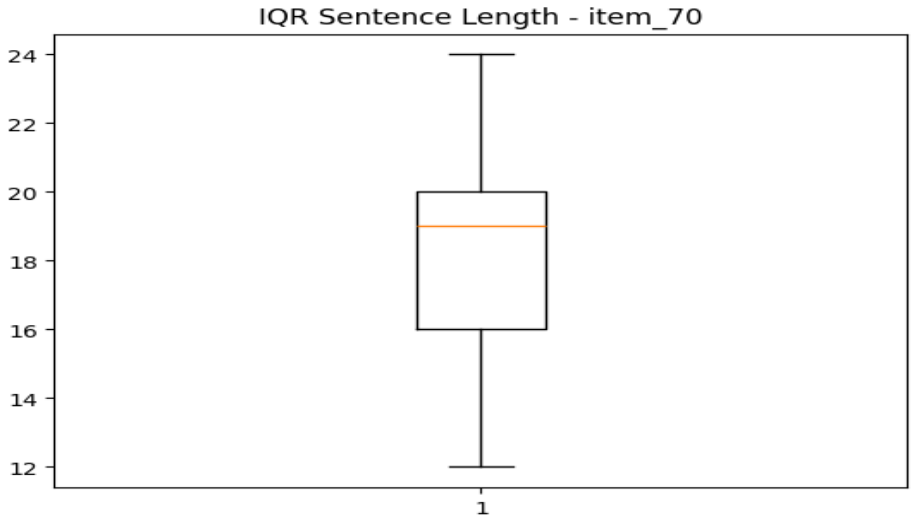
sentence_IQR_item_69.png



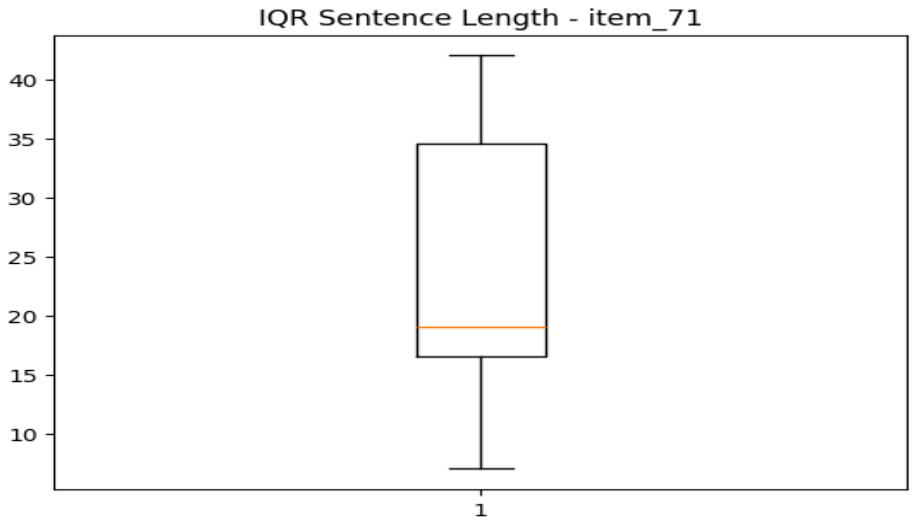
sentence_IQR_item_7.png



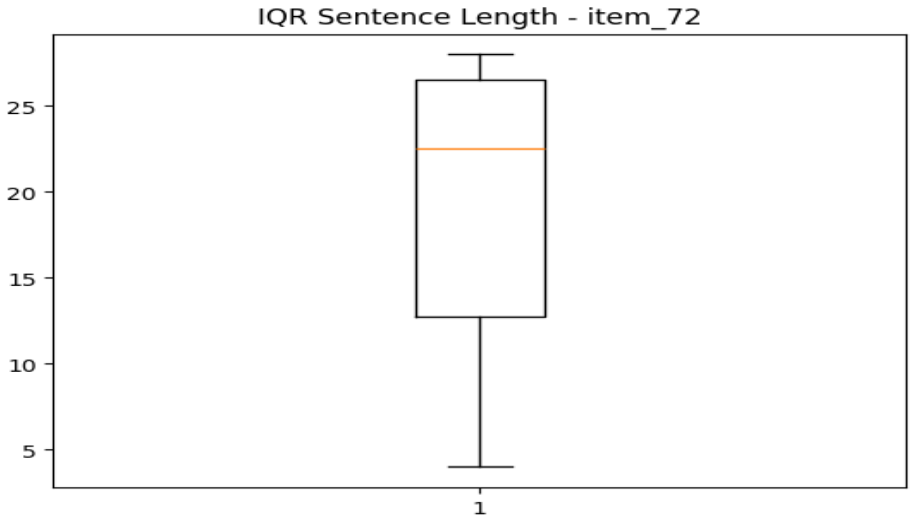
sentence_IQR_item_70.png



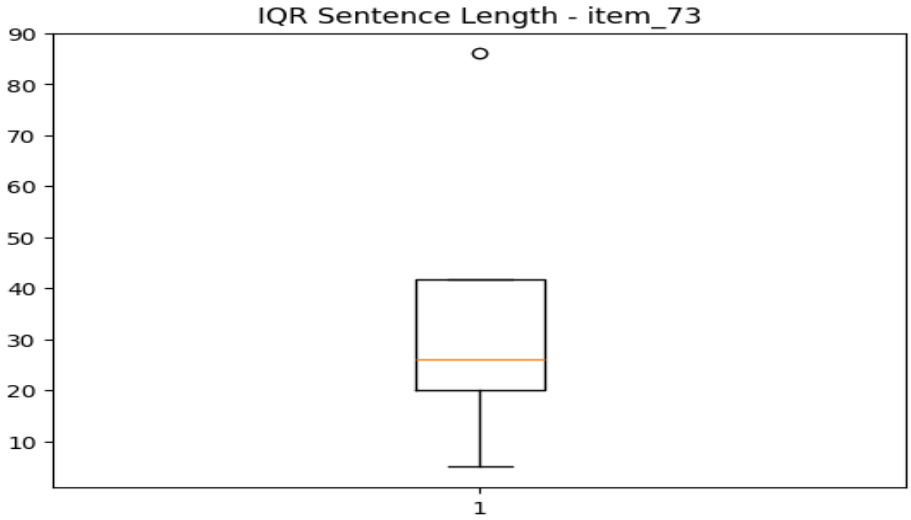
sentence_IQR_item_71.png



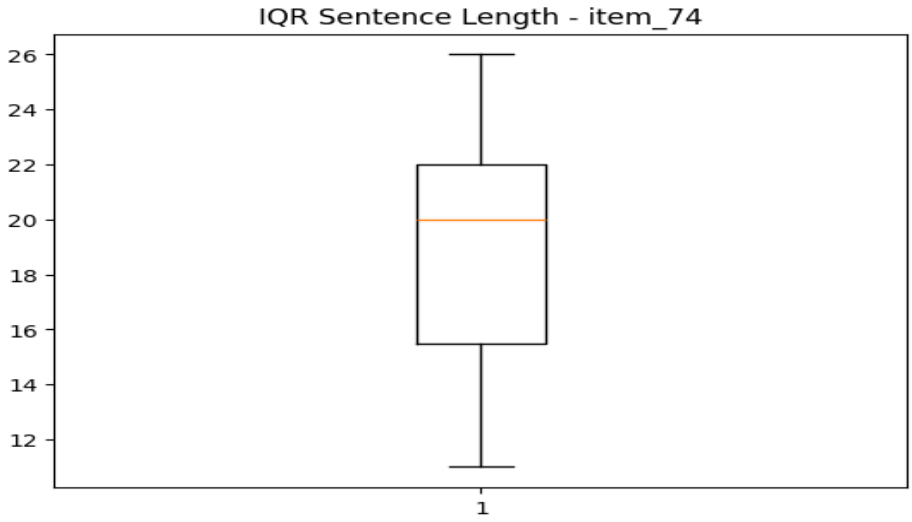
sentence_IQR_item_72.png



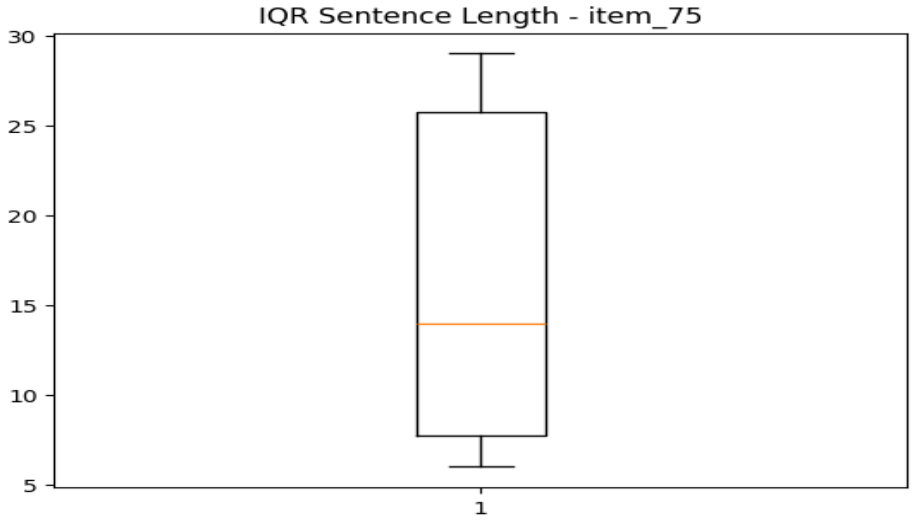
sentence_IQR_item_73.png



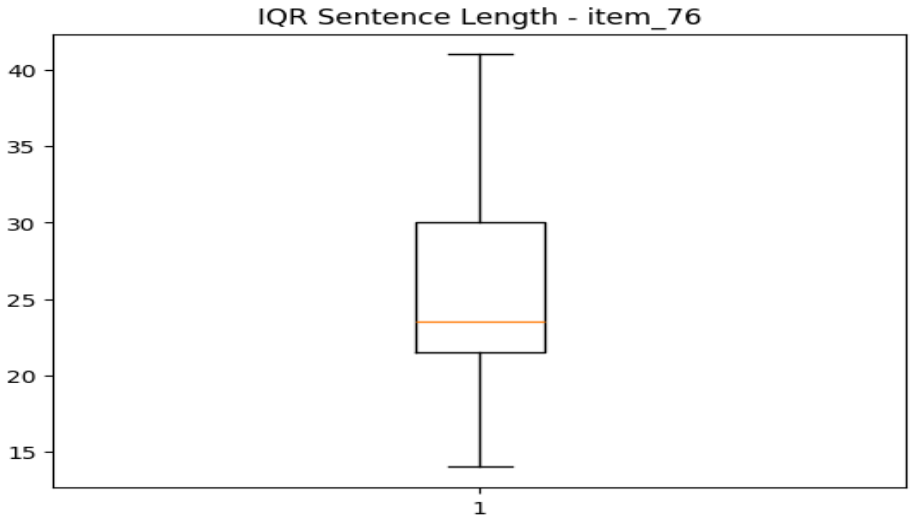
sentence_IQR_item_74.png



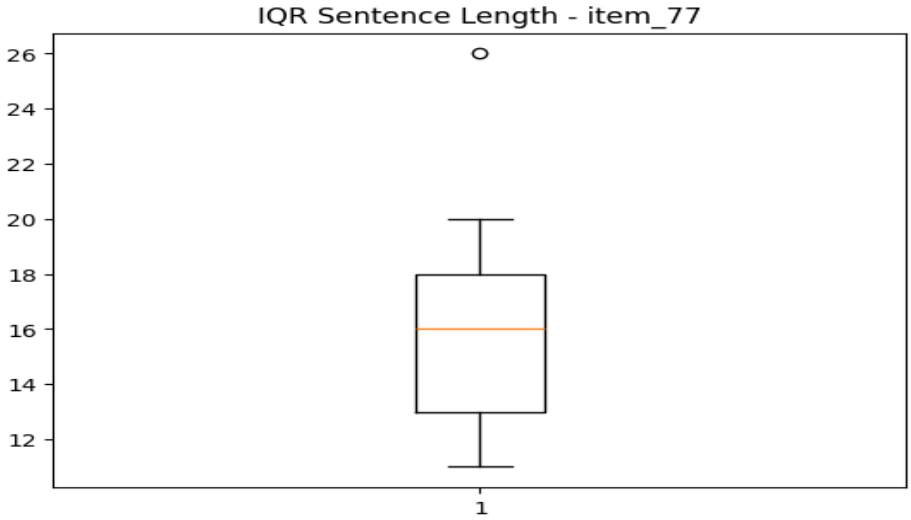
sentence_IQR_item_75.png



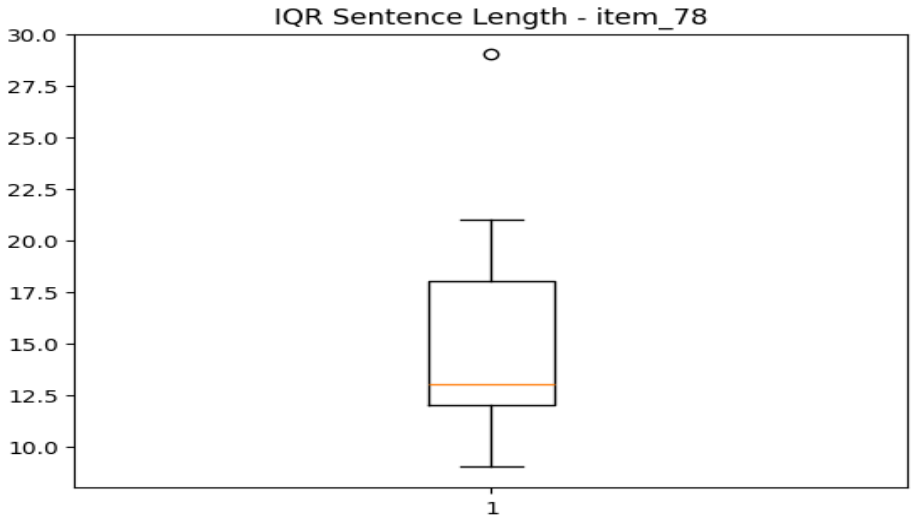
sentence_IQR_item_76.png



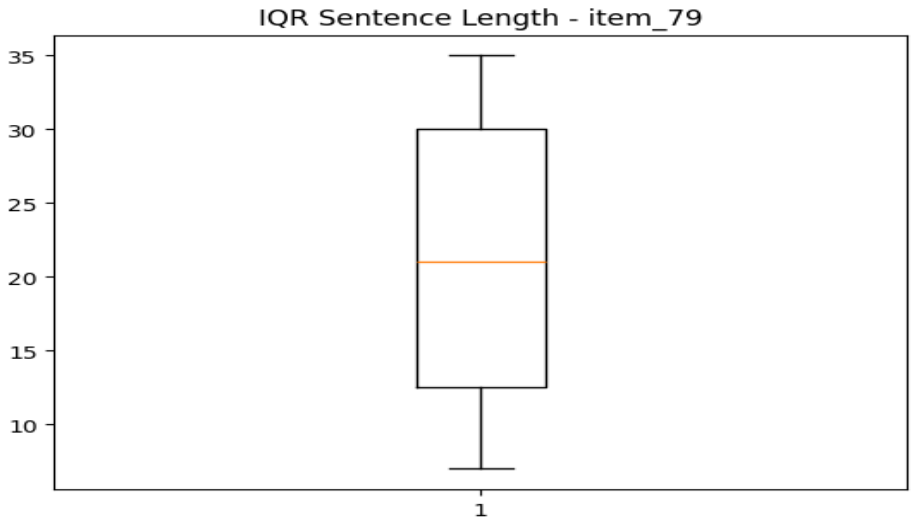
sentence_IQR_item_77.png



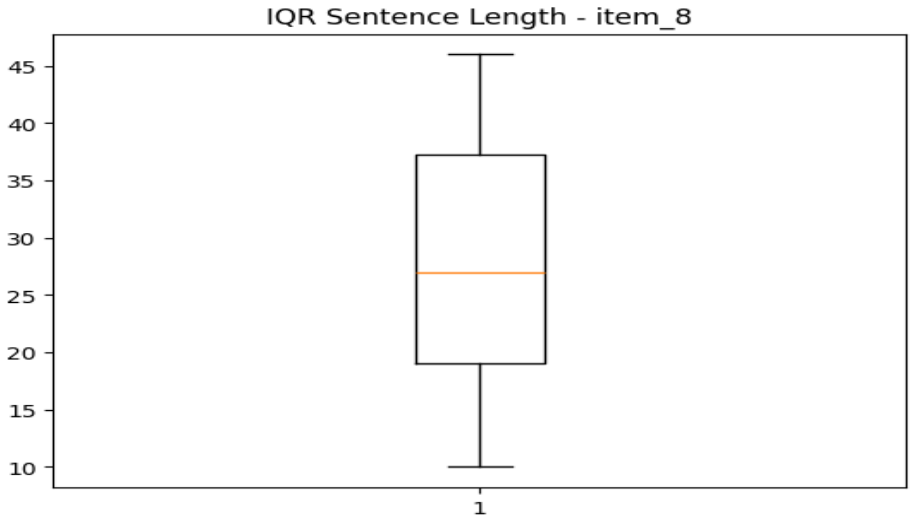
sentence_IQR_item_78.png



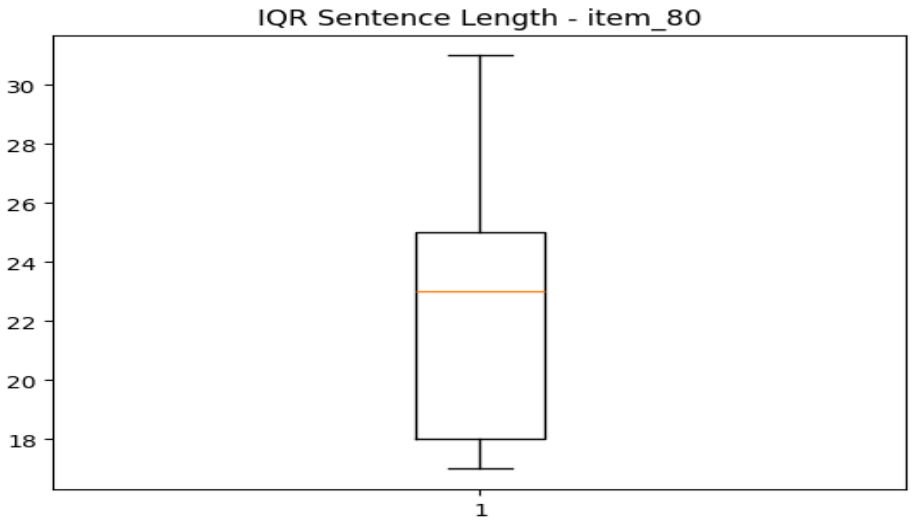
sentence_IQR_item_79.png



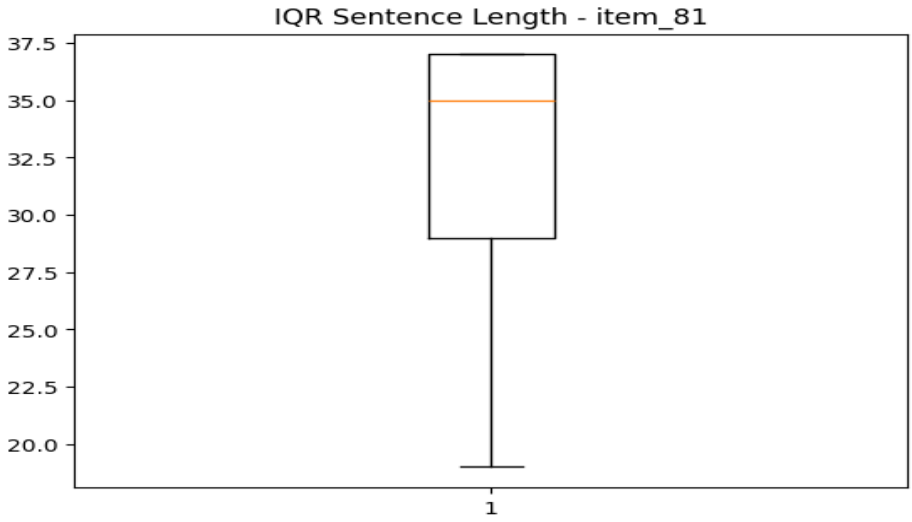
sentence_IQR_item_8.png



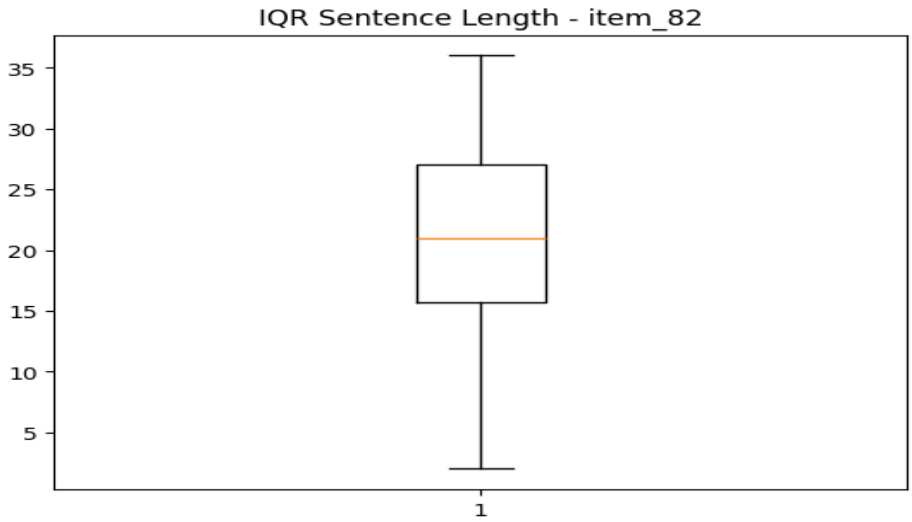
sentence_IQR_item_80.png



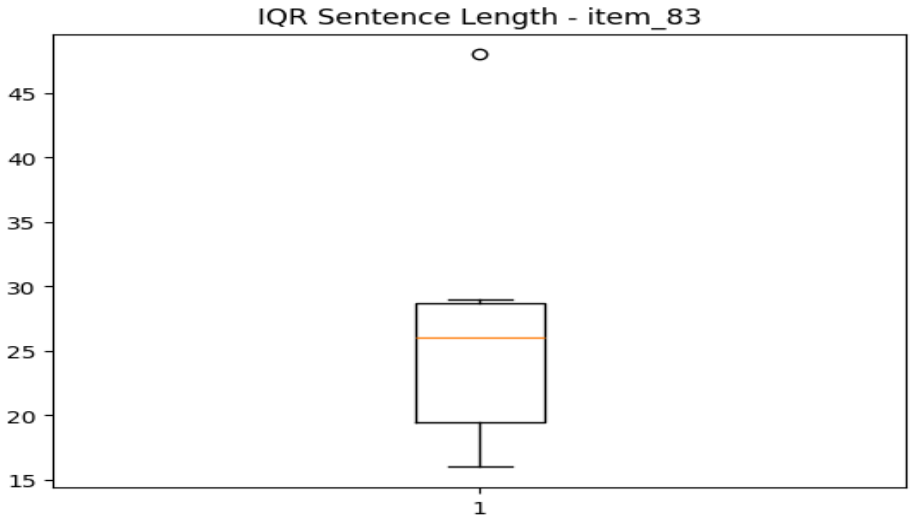
sentence_IQR_item_81.png



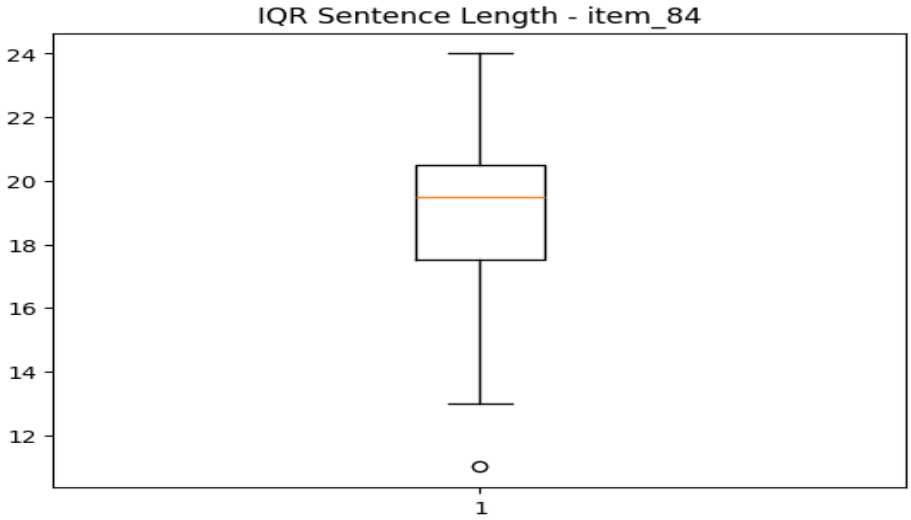
sentence_IQR_item_82.png



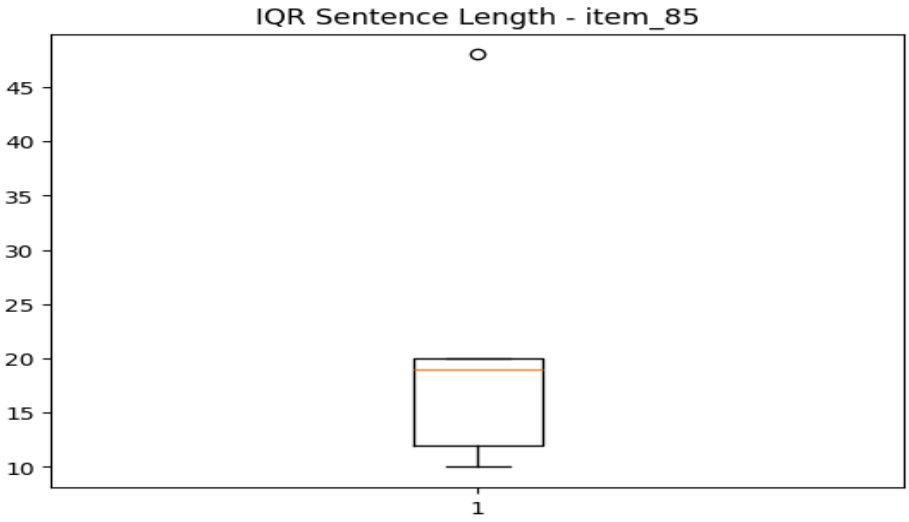
sentence_IQR_item_83.png



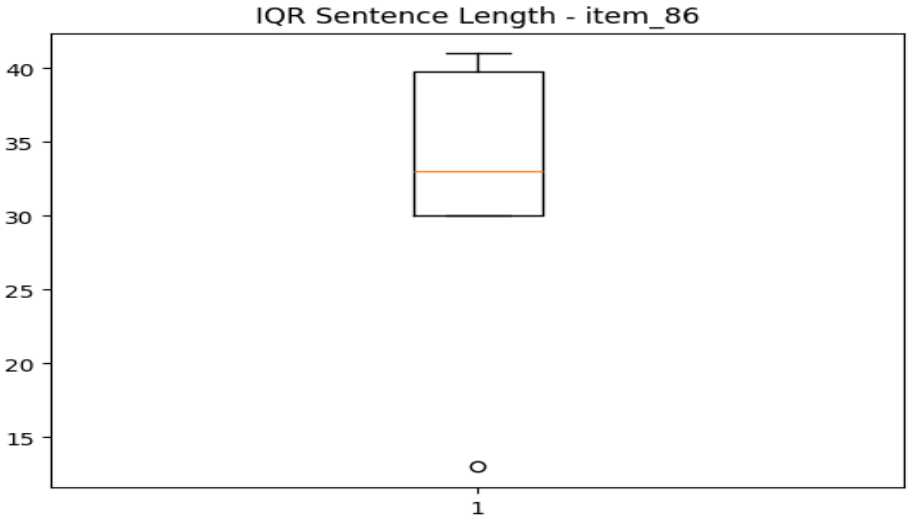
sentence_IQR_item_84.png



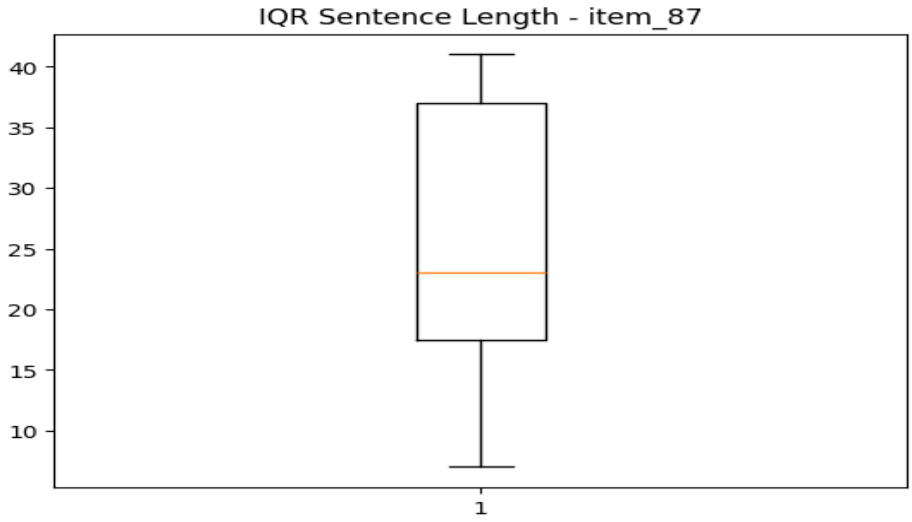
sentence_IQR_item_85.png



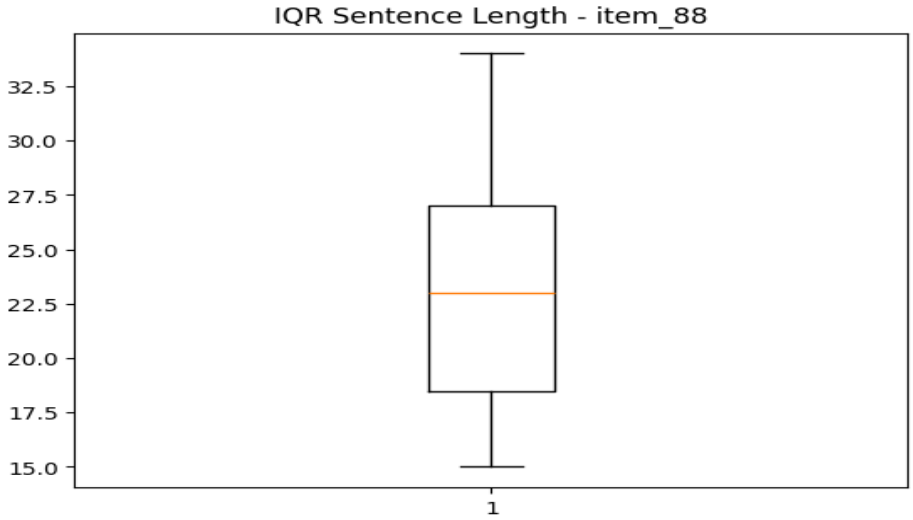
sentence_IQR_item_86.png



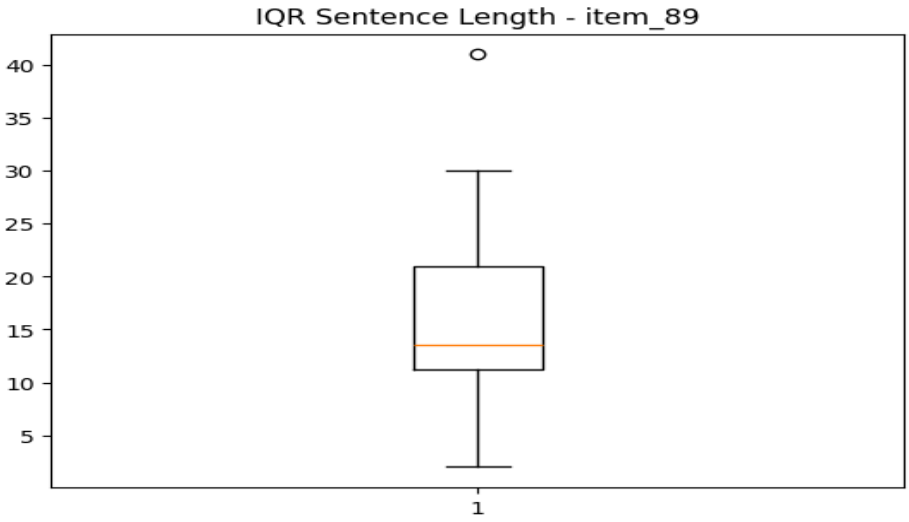
sentence_IQR_item_87.png



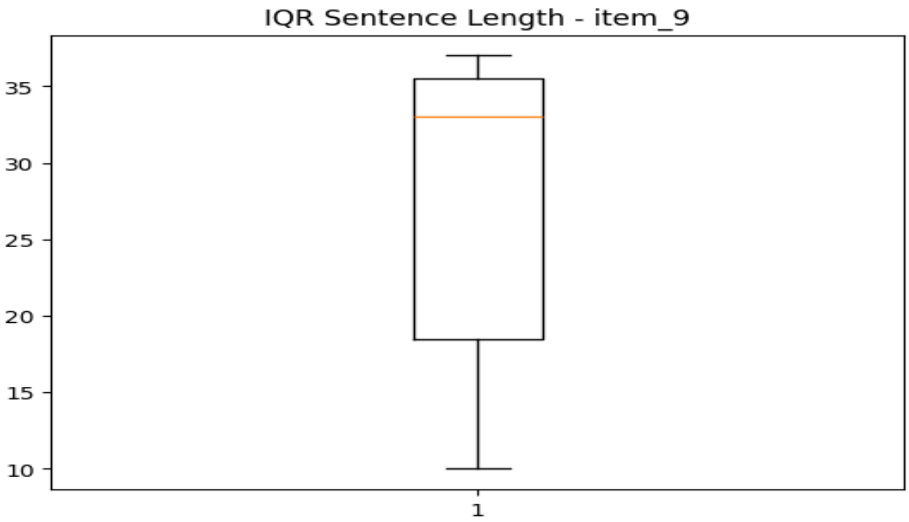
sentence_IQR_item_88.png



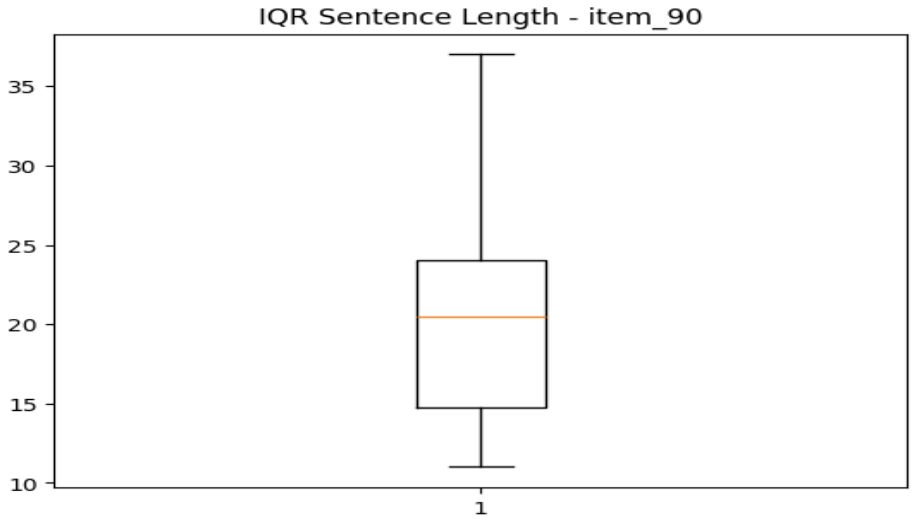
sentence_IQR_item_89.png



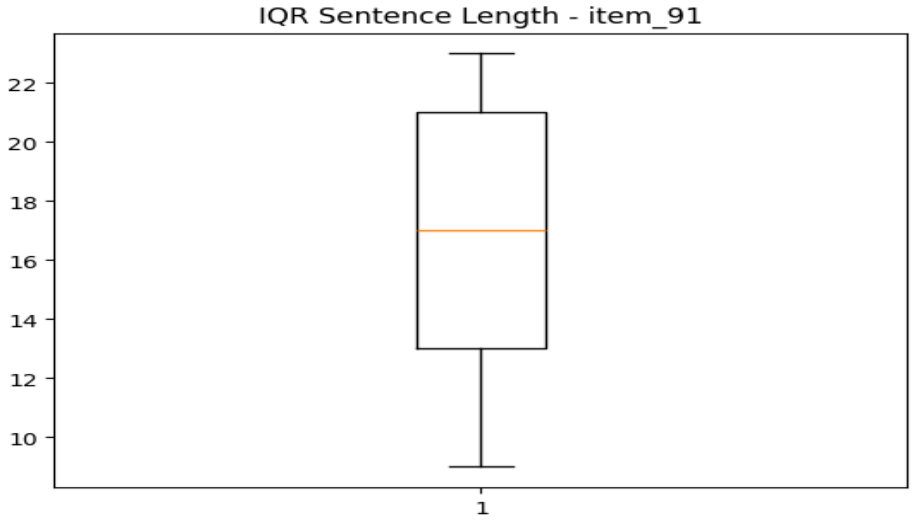
sentence_IQR_item_9.png



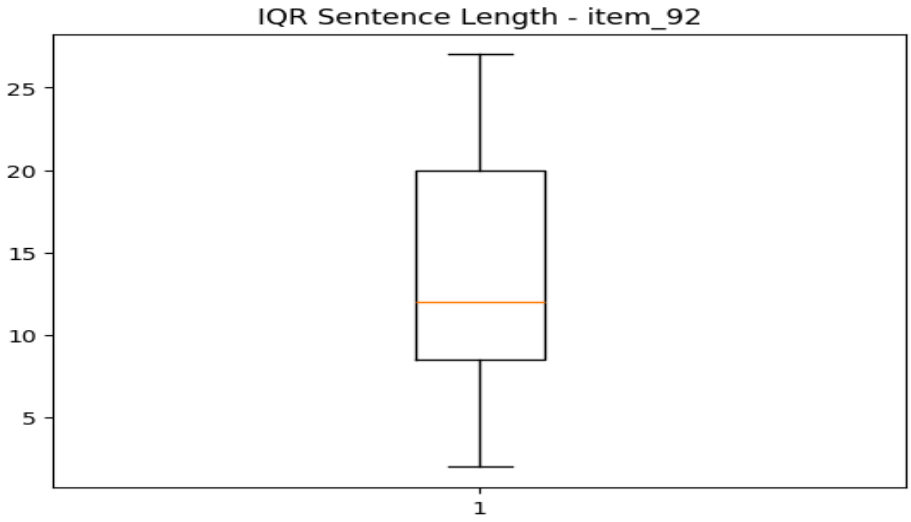
sentence_IQR_item_90.png



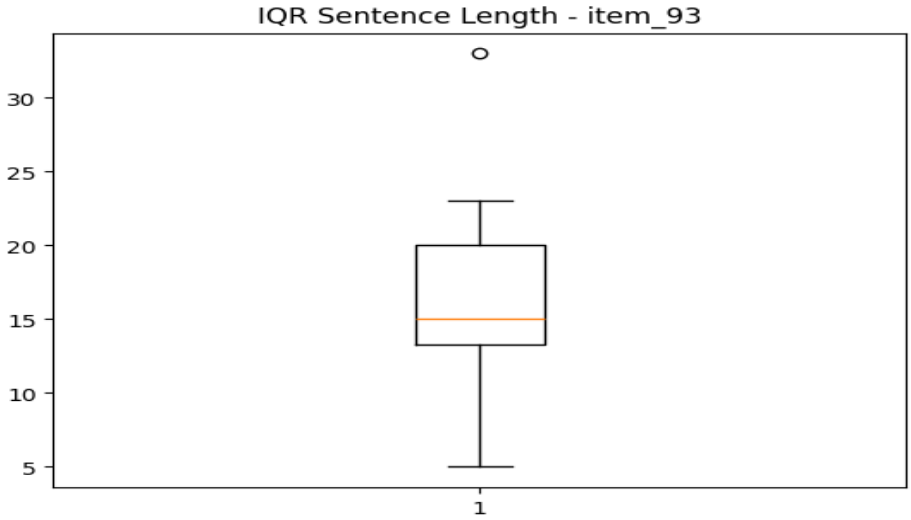
sentence_IQR_item_91.png



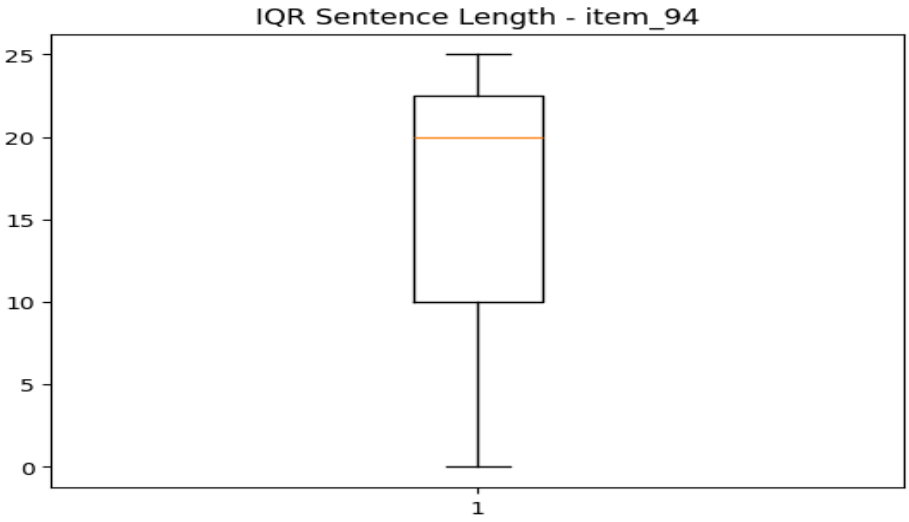
sentence_IQR_item_92.png



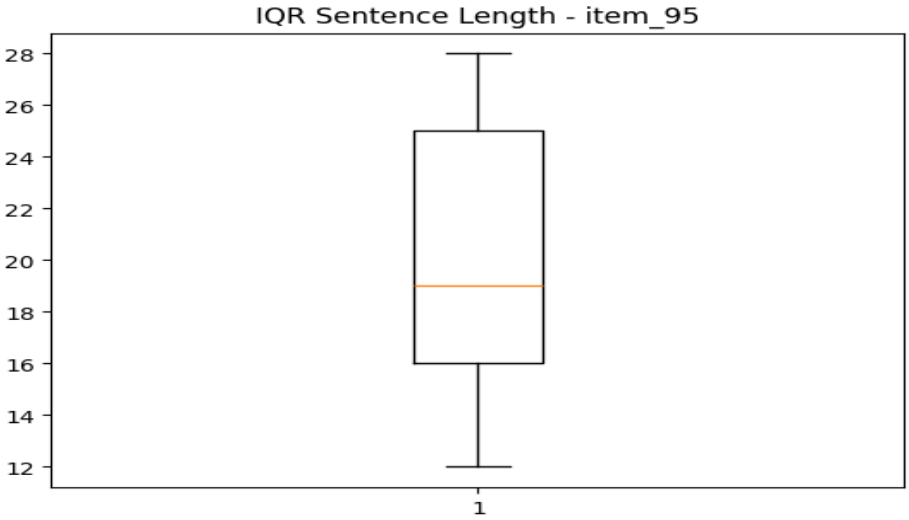
sentence_IQR_item_93.png



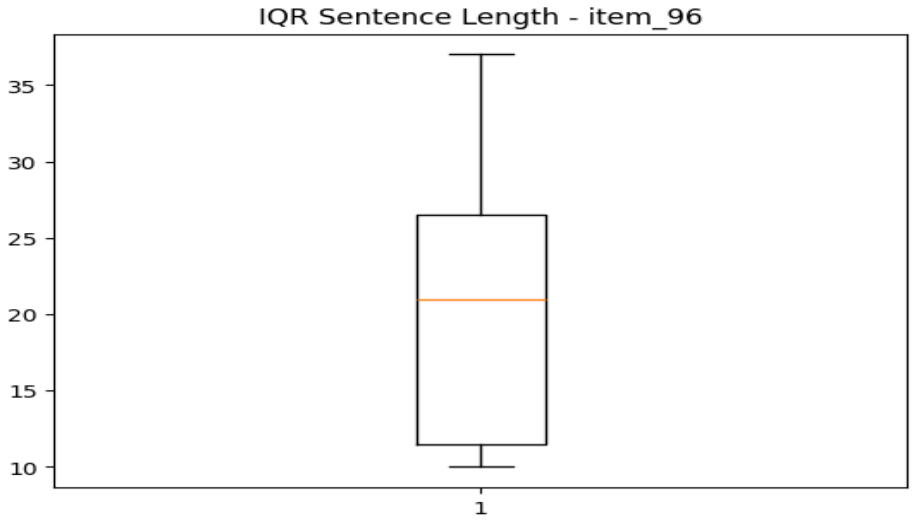
sentence_IQR_item_94.png



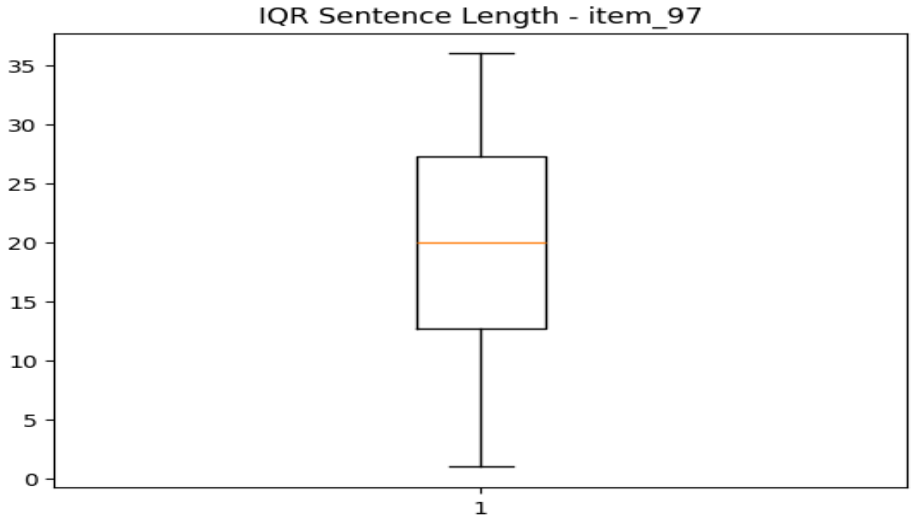
sentence_IQR_item_95.png



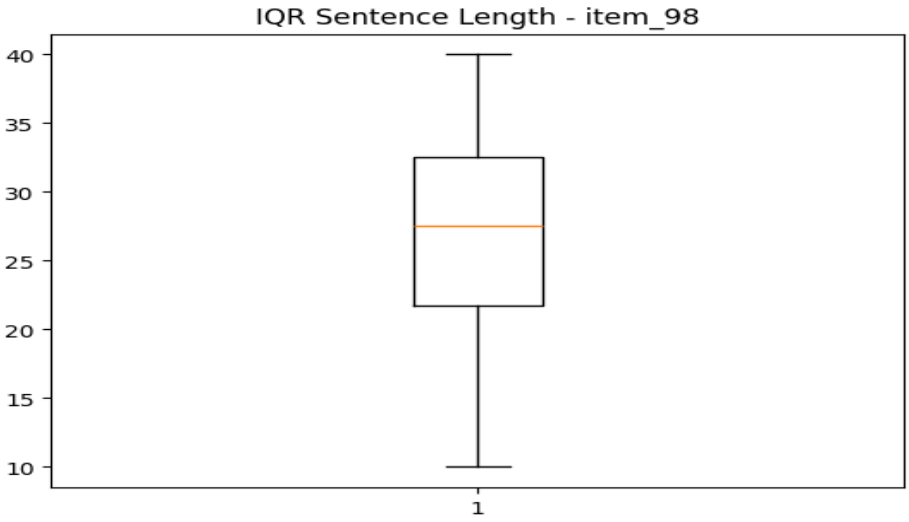
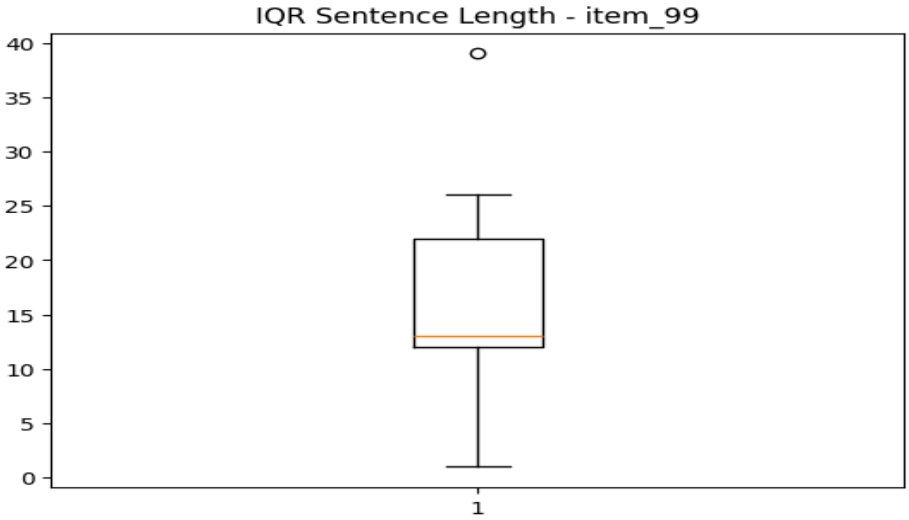
sentence_IQR_item_96.png



sentence_IQR_item_97.png



sentence_IQR_item_98.png



sentence_IQR_item_99.png